Learning from data using probability

Carlos Fernandez-Granda
www.cims.nyu.edu/~cfgranda

cSplash 2019

Acknowledgements

Support by NSF award DMS-1616340

Voting in the House of Representatives (1984)

Affiliation	Vote 1	Vote 2	Vote 3	...	Vote 16
Republican	No	Yes	No	...	Yes
Democrat	Yes	Yes	Yes	...	–
Republican	No	–	No	...	Yes
Democrat	No	Yes	Yes	...	No
Democrat	Yes	Yes	Yes	...	No
???	Yes	Yes	–	...	No
???	No	Yes	No	...	No

Probabilistic modeling

Probability enables us to quantify uncertainty

How likely is someone to be a democrat?

How likely is someone to vote Yes in Vote 3?

How likely is someone to vote Yes in Vote 3 if they are a republican?

Challenges

How to estimate probabilities from data

How to combine them to make predictions

Random variable

Mathematical objects that model uncertain quantities

A random variable X has a set of possible outcomes

Examples

- **Affiliation**. Outcomes: Democrat or Republican
- **Vote 1**. Outcomes: Yes or No

Probability

Maps outcomes to a number between 0 and 1

The probability of an outcome quantifies how likely it is

Estimating probabilities

$$\mathrm{P}\left(\text{Affiliation} = \text{Democrat}\right) = \rule{4cm}{0.4pt}$$

Estimating probabilities

$$\mathrm{P}\left(\text{Affiliation} = \text{Democrat}\right) = \frac{\#\text{Democrats}}{}$$

Estimating probabilities

$$\mathrm{P}\left(\text{Affiliation} = \text{Democrat}\right) = \frac{\#\text{Democrats}}{\text{Total}}$$

Estimating probabilities

$$\mathrm{P}\left(\text{Affiliation} = \text{Democrat}\right) = \frac{\#\text{Democrats}}{\text{Total}} = \frac{267}{435} = 0.614$$

Estimating probabilities

$$\mathrm{P}\left(\text{Vote 1} = \text{Yes}\right) = \underline{\hspace{2cm}}$$

Estimating probabilities

$$\mathrm{P}\left(\text{Vote 1} = \text{Yes}\right) = \frac{\#\text{Yes}}{}$$

Estimating probabilities

$$\mathrm{P}\left(\text{Vote 1} = \text{Yes}\right) = \frac{\#\text{Yes}}{\text{Total}}$$

Estimating probabilities

$$\mathrm{P}\left(\text{Vote 1} = \text{Yes}\right) = \frac{\#\text{Yes}}{\text{Total}}$$
$$= \frac{187}{423} = 0.442$$

Properties of probability

Probability is nonnegative, like mass or length

Properties of probability

The probability of all possible outcomes adds to one

$$\mathrm{P}\left(\text{Vote 1 = Yes}\right) + \mathrm{P}\left(\text{Vote 1 = No}\right) = \_\_\_\_\_ + \_\_\_\_\_$$

Properties of probability

The probability of all possible outcomes adds to one

$$\mathrm{P}\left(\text{Vote 1 = Yes}\right) + \mathrm{P}\left(\text{Vote 1 = No}\right) = \frac{\#\text{Yes}}{\text{Total}} + \underline{\hspace{3em}}$$

Properties of probability

The probability of all possible outcomes adds to one

$$\mathrm{P}\left(\text{Vote 1} = \text{Yes}\right) + \mathrm{P}\left(\text{Vote 1} = \text{No}\right) = \frac{\#\text{Yes}}{\text{Total}} + \frac{\#\text{No}}{\text{Total}}$$

Properties of probability

The probability of all possible outcomes adds to one

$$
\begin{aligned}
\mathrm{P}(\text{Vote 1} = \text{Yes}) + \mathrm{P}(\text{Vote 1} = \text{No}) &= \frac{\#\text{Yes}}{\text{Total}} + \frac{\#\text{No}}{\text{Total}} \\
&= \frac{\text{Total}}{\text{Total}}
\end{aligned}
$$

Properties of probability

The probability of all possible outcomes adds to one

$$\mathrm{P}\left(\text{Vote 1 = Yes}\right) + \mathrm{P}\left(\text{Vote 1 = No}\right) = \frac{\#\text{Yes}}{\text{Total}} + \frac{\#\text{No}}{\text{Total}}$$
$$= \frac{\text{Total}}{\text{Total}}$$
$$= 1$$

Properties of probability

The probability of all possible outcomes adds to one

$$\begin{aligned} \mathrm{P}\left(\text{Vote 1 = Yes}\right) + \mathrm{P}\left(\text{Vote 1 = No}\right) &= \frac{\#\text{Yes}}{\text{Total}} + \frac{\#\text{No}}{\text{Total}} \\ &= \frac{\text{Total}}{\text{Total}} \\ &= 1 \end{aligned}$$

Not like mass or length!

Multiple random variables

We can consider several random variables at the same time

$$\mathrm{P}\left(\text{Affiliation} = \text{R and Vote 1} = \text{Yes}\right) = \underline{\hspace{3cm}}$$

Multiple random variables

We can consider several random variables at the same time

$$\mathrm{P}\left(\text{Affiliation} = \text{R and Vote } 1 = \text{Yes}\right) = \frac{\#\text{R and Yes}}{}$$

Multiple random variables

We can consider several random variables at the same time

$$\mathrm{P}\left(\text{Affiliation} = \text{R and Vote 1} = \text{Yes}\right) = \frac{\#\text{R and Yes}}{\text{Total}}$$

Multiple random variables

We can consider several random variables at the same time

$$\mathrm{P}\left(\text{Affiliation} = \text{R and Vote 1} = \text{Yes}\right) = \frac{\#\text{R and Yes}}{\text{Total}}$$
$$= \frac{31}{423} = 0.073$$

Conditional probability

Quantifies uncertainty if we have partial information

$$\mathrm{P}\left(\text{Vote 1} = \text{Yes} \;\middle|\; \text{Affiliation} = \text{R}\right) = \rule{3cm}{0.4pt}$$

Conditional probability

Quantifies uncertainty if we have partial information

$$\mathrm{P}\left(\text{Vote 1} = \text{Yes} \;\mid\; \text{Affiliation} = \text{R}\right) = \frac{\#\text{Yes and R}}{}$$

Conditional probability

Quantifies uncertainty if we have partial information

$$\mathrm{P}\left(\text{Vote 1} = \text{Yes} \;\middle|\; \text{Affiliation} = \text{R}\right) = \frac{\#\text{Yes and R}}{\#\ \text{R}}$$

Conditional probability

Quantifies uncertainty if we have partial information

$$\mathrm{P}\left(\text{Vote 1} = \text{Yes} \;\middle|\; \text{Affiliation} = \text{R}\right) = \frac{\#\text{Yes and R}}{\#\ \text{R}}$$

$$= \frac{31}{168} = 0.185$$

Chain rule

$$\mathrm{P}\left(X = A \text{ and } Y = B\right) = \mathrm{P}\left(X = A\right) \mathrm{P}\left(Y = B \,|\, X = A\right)$$

Chain rule

$$\mathrm{P}(X = A \text{ and } Y = B) = \mathrm{P}(X = A)\,\mathrm{P}(Y = B \mid X = A)$$

Proof:

$$\mathrm{P}(X = A)\,\mathrm{P}(Y = B \mid X = A) =$$

Chain rule

$$\mathrm{P}\left(X = A \text{ and } Y = B\right) = \mathrm{P}\left(X = A\right) \mathrm{P}\left(Y = B \,|\, X = A\right)$$

Proof:

$$\mathrm{P}\left(X = A\right) \mathrm{P}\left(Y = B \,|\, X = A\right) = \frac{\# X = A}{\text{Total}} \cdot$$

Chain rule

$$\mathrm{P}\left(X = A \text{ and } Y = B\right) = \mathrm{P}\left(X = A\right)\mathrm{P}\left(Y = B \,|\, X = A\right)$$

Proof:

$$\mathrm{P}\left(X = A\right)\mathrm{P}\left(Y = B \,|\, X = A\right) = \frac{\# X = A}{\text{Total}} \cdot \frac{\# \; Y = B \text{ and } X = A}{\# X = A}$$

Chain rule

$$\mathrm{P}\left(X = A \text{ and } Y = B\right) = \mathrm{P}\left(X = A\right)\mathrm{P}\left(Y = B \,|\, X = A\right)$$

Proof:

$$\begin{aligned}\mathrm{P}\left(X = A\right)\mathrm{P}\left(Y = B \,|\, X = A\right) &= \frac{\# X = A}{\text{Total}} \cdot \frac{\#\ Y = B \text{ and } X = A}{\# X = A} \\ &= \frac{\#\ Y = B \text{ and } X = A}{\text{Total}}\end{aligned}$$

Chain rule

$$\mathrm{P}\left(X = A \text{ and } Y = B\right) = \mathrm{P}\left(X = A\right)\mathrm{P}\left(Y = B \,|\, X = A\right)$$

Proof:

$$\begin{aligned}\mathrm{P}\left(X = A\right)\mathrm{P}\left(Y = B \,|\, X = A\right) &= \frac{\# X = A}{\text{Total}} \cdot \frac{\# \; Y = B \text{ and } X = A}{\# X = A} \\ &= \frac{\# \; Y = B \text{ and } X = A}{\text{Total}} \\ &= \mathrm{P}\left(X = A \text{ and } Y = B\right)\end{aligned}$$

Predicting affiliation

Affiliation	Vote 1	Vote 2	Vote 3	...	Vote 16
Republican	No	Yes	No	...	Yes
Democrat	Yes	Yes	Yes	...	–
Republican	No	–	No	...	Yes
Democrat	No	Yes	Yes	...	No
Democrat	Yes	Yes	Yes	...	No
???	Yes	Yes	–	...	No
???	No	Yes	No	...	No

Predicting affiliation

Goal: Compute probability of Affiliation conditioned on Votes

By the chain rule

$$\mathrm{P}\left(\text{Aff} = \text{R} \mid \text{V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N}\right)$$
$$= \frac{\mathrm{P}\left(\text{Aff} = \text{R and V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N}\right)}{\mathrm{P}\left(\text{V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N}\right)}$$

Predicting affiliation

Goal: Compute probability of Affiliation conditioned on Votes

By the chain rule

$$\mathrm{P}\left(\text{Aff} = \text{R} \mid \text{V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N}\right) = \frac{\mathrm{P}\left(\text{Aff} = \text{R and V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N}\right)}{\mathrm{P}\left(\text{V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N}\right)}$$

Problem: How do we estimate these probabilities?

Predicting affiliation

$$\mathrm{P}\left(\mathsf{V1} = \mathsf{Y} \text{ and } \mathsf{V2} = \mathsf{Y} \text{ and } \ldots \mathsf{V16} = \mathsf{N}\right)$$

$$= \frac{\qquad\qquad}{\qquad\qquad}$$

Predicting affiliation

$$\mathrm{P}\left(\text{V1 = Y and V2 = Y and} \ldots \text{V16 = N}\right) = \frac{\#\text{V1 = Y and V2 = Y and} \ldots \text{V16 = N}}{\text{Total}}$$

Predicting affiliation

$$\mathrm{P}\left(\text{V1 = Y and V2 = Y and} \ldots \text{V16 = N}\right) = \frac{\#\text{V1 = Y and V2 = Y and} \ldots \text{V16 = N}}{\text{Total}}$$

Problem: Many different possibilities!

Predicting affiliation

$$\mathrm{P}\left(\text{V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N}\right)$$
$$= \frac{\#\text{V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N}}{\text{Total}}$$

Problem: Many different possibilities! ($2^{16} = 65{,}536$)

Predicting affiliation

$$\mathrm{P}\left(\text{V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N}\right)$$
$$= \frac{\#\text{V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N}}{\text{Total}}$$

Problem: Many different possibilities! ($2^{16} = 65{,}536$)

There's only 435 politicians...

Predicting affiliation

$$\mathrm{P}\left(\text{V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N}\right)$$
$$= \frac{\#\text{V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N}}{\text{Total}}$$

Problem: Many different possibilities! ($2^{16} = 65{,}536$)

There's only 435 politicians...

Most politicians have unique sequence of votes (304 out of 435)

Predicting affiliation

$$
\begin{aligned}
&\mathrm{P}\left(\text{V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N}\right) \\
&= \frac{\#\text{V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N}}{\text{Total}} \\
&= \text{ mostly } 0 \ldots
\end{aligned}
$$

Problem: Many different possibilities! ($2^{16} = 65{,}536$)

There's only 435 politicians...

Most politicians have unique sequence of votes (304 out of 435)

Independence

If knowing that $X = A$ happened does not affect how likely it is that $Y = B$ then X and Y are independent

$$\mathrm{P}\left(Y = B \mid X = A\right) = \mathrm{P}\left(Y = B\right)$$

Independence

If knowing that $X = A$ happened does not affect how likely it is that $Y = B$ then X and Y are independent

$$\mathrm{P}\left(Y = B \,|\, X = A\right) = \mathrm{P}\left(Y = B\right)$$

In that case

$$\mathrm{P}\left(X = A \text{ and } Y = B\right)$$

Independence

If knowing that $X = A$ happened does not affect how likely it is that $Y = B$ then X and Y are independent

$$\mathrm{P}\left(Y = B \,|\, X = A\right) = \mathrm{P}\left(Y = B\right)$$

In that case

$$\mathrm{P}\left(X = A \text{ and } Y = B\right) = \mathrm{P}\left(X = A\right)\mathrm{P}\left(Y = B \,|\, X = A\right)$$

Independence

If knowing that $X = A$ happened does not affect how likely it is that $Y = B$ then X and Y are independent

$$\mathrm{P}(Y = B \,|\, X = A) = \mathrm{P}(Y = B)$$

In that case

$$\begin{aligned}\mathrm{P}(X = A \text{ and } Y = B) &= \mathrm{P}(X = A)\,\mathrm{P}(Y = B \,|\, X = A) \\ &= \mathrm{P}(X = A)\,\mathrm{P}(Y = B)\end{aligned}$$

Independence

If votes are independent

$$\begin{aligned} &\mathrm{P}(\mathsf{V1} = \mathsf{Y} \text{ and } \mathsf{V2} = \mathsf{Y} \text{ and } \ldots \mathsf{V16} = \mathsf{N}) \\ &\quad = \mathrm{P}(\mathsf{V1} = \mathsf{Y})\,\mathrm{P}(\mathsf{V2} = \mathsf{Y}) \ldots \mathrm{P}(\mathsf{V16} = \mathsf{N}) \end{aligned}$$

Are votes independent?

$$\mathrm{P}\left(\text{Vote 4= Yes}\right) = 0.505$$

$$\mathrm{P}\left(\text{Vote 11 = Yes}\right) = 0.423$$

$$\mathrm{P}\left(\text{Vote 4= Yes and Vote 11 = Yes}\right) = 0.378$$

Are votes independent?

$$\mathrm{P}\left(\text{Vote 4= Yes}\right) = 0.505$$

$$\mathrm{P}\left(\text{Vote 11 = Yes}\right) = 0.423$$

$$\mathrm{P}\left(\text{Vote 4= Yes and Vote 11 = Yes}\right) = 0.378$$

$$\mathrm{P}\left(\text{Vote 4= Yes}\right)\mathrm{P}\left(\text{Vote 11 = Yes}\right) =$$

Are votes independent?

$$\mathrm{P}\left(\text{Vote 4= Yes}\right) = 0.505$$

$$\mathrm{P}\left(\text{Vote 11 = Yes}\right) = 0.423$$

$$\mathrm{P}\left(\text{Vote 4= Yes and Vote 11 = Yes}\right) = 0.378$$

$$\mathrm{P}\left(\text{Vote 4= Yes}\right)\mathrm{P}\left(\text{Vote 11 = Yes}\right) = 0.214$$

Chain rule

$$\mathrm{P}\left(\text{X=A and Y = B} \mid \text{Z = C}\right) = \mathrm{P}\left(X = A \,|\, Z = C\right) \mathrm{P}\left(\text{Y = B} \mid \text{X=A and Z=C}\right)$$

Chain rule

$$\mathrm{P}\left(\text{X=A and Y = B} \mid \text{Z = C}\right) = \mathrm{P}\left(X = A \,|\, Z = C\right) \mathrm{P}\left(\text{Y = B} \mid \text{X=A and Z=C}\right)$$

Proof:

$$\mathrm{P}\left(X = A \,|\, Z = C\right) \mathrm{P}\left(\text{Y = B} \mid \text{X=A and Z=C}\right)$$

$$=$$

Chain rule

$$\mathrm{P}\left(\text{X=A and Y = B} \mid \text{Z = C}\right) = \mathrm{P}\left(X = A \,|\, Z = C\right) \mathrm{P}\left(\text{Y = B} \mid \text{X=A and Z=C}\right)$$

Proof:

$$\begin{aligned} &\mathrm{P}\left(X = A \,|\, Z = C\right) \mathrm{P}\left(\text{Y = B} \mid \text{X=A and Z=C}\right) \\ &\quad = \frac{\#X = A \text{ and } Z = C}{\#Z = C} \cdot \end{aligned}$$

Chain rule

$$\mathrm{P}\left(\text{X=A and Y = B} \mid \text{Z = C}\right) = \mathrm{P}\left(X = A \,|\, Z = C\right) \mathrm{P}\left(\text{Y = B} \mid \text{X=A and Z=C}\right)$$

Proof:

$$\begin{aligned} &\mathrm{P}\left(X = A \,|\, Z = C\right) \mathrm{P}\left(\text{Y = B} \mid \text{X=A and Z=C}\right) \\ &= \frac{\#X = A \text{ and } Z = C}{\#Z = C} \cdot \frac{\#\ Y = B \text{ and } X = A \text{ and } Z = C}{\#X = A \text{ and } Z = C} \end{aligned}$$

Chain rule

$$\mathrm{P}\left(\text{X=A and Y = B} \mid \text{Z = C}\right) = \mathrm{P}\left(X = A \,|\, Z = C\right)\mathrm{P}\left(\text{Y = B} \mid \text{X=A and Z=C}\right)$$

Proof:

$$\begin{aligned}
&\mathrm{P}\left(X = A \,|\, Z = C\right)\mathrm{P}\left(\text{Y = B} \mid \text{X=A and Z=C}\right) \\
&\quad = \frac{\#X = A \text{ and } Z = C}{\#Z = C} \cdot \frac{\#\ Y = B \text{ and } X = A \text{ and } Z = C}{\#X = A \text{ and } Z = C} \\
&\quad = \frac{\#\ Y = B \text{ and } X = A \text{ and } Z = C}{\#Z = C}
\end{aligned}$$

Chain rule

$$\mathrm{P}(\text{X=A and Y = B} \mid \text{Z = C}) = \mathrm{P}(X = A \mid Z = C)\,\mathrm{P}(\text{Y = B} \mid \text{X=A and Z=C})$$

Proof:

$$\begin{aligned}
&\mathrm{P}(X = A \mid Z = C)\,\mathrm{P}(\text{Y = B} \mid \text{X=A and Z=C}) \\
&\quad = \frac{\#X = A \text{ and } Z = C}{\#Z = C} \cdot \frac{\#\ Y = B \text{ and } X = A \text{ and } Z = C}{\#X = A \text{ and } Z = C} \\
&\quad = \frac{\#\ Y = B \text{ and } X = A \text{ and } Z = C}{\#Z = C} \\
&\quad = \mathrm{P}(\text{X=A and Y = B} \mid \text{Z = C})
\end{aligned}$$

Conditional independence

If knowing that $X = A$ happened does not affect how likely it is that $Y = B$ if $Z = C$, then X and Y are independent conditioned on $Z = C$

$$\mathrm{P}\left(Y = B \,|\, X = A \text{ and } Z = C\right) = \mathrm{P}\left(Y = B \,|\, Z = C\right)$$

Conditional independence

If knowing that $X = A$ happened does not affect how likely it is that $Y = B$ if $Z = C$, then X and Y are independent conditioned on $Z = C$

$$\mathrm{P}\left(Y = B \,|\, X = A \text{ and } Z = C\right) = \mathrm{P}\left(Y = B \,|\, Z = C\right)$$

In that case

Conditional independence

If knowing that $X = A$ happened does not affect how likely it is that $Y = B$ if $Z = C$, then X and Y are independent conditioned on $Z = C$

$$\mathrm{P}\left(Y = B \,|\, X = A \text{ and } Z = C\right) = \mathrm{P}\left(Y = B \,|\, Z = C\right)$$

In that case

P (X=A and Y=B | Z=C)

Conditional independence

If knowing that $X = A$ happened does not affect how likely it is that $Y = B$ if $Z = C$, then X and Y are independent conditioned on $Z = C$

$$\mathrm{P}\left(Y = B \,|\, X = A \text{ and } Z = C\right) = \mathrm{P}\left(Y = B \,|\, Z = C\right)$$

In that case

$$\mathrm{P}\left(\text{X=A and Y=B} \mid \text{Z=C}\right) = \mathrm{P}\left(X = A \,|\, Z = C\right)\mathrm{P}\left(\text{Y=B} \mid \text{X=A and Z=C}\right)$$

Conditional independence

If knowing that $X = A$ happened does not affect how likely it is that $Y = B$ if $Z = C$, then X and Y are independent conditioned on $Z = C$

$$\mathrm{P}\left(Y = B \,|\, X = A \text{ and } Z = C\right) = \mathrm{P}\left(Y = B \,|\, Z = C\right)$$

In that case

$$\begin{aligned} \mathrm{P}\left(\text{X=A and Y=B} \mid \text{Z=C}\right) &= \mathrm{P}\left(X = A \,|\, Z = C\right) \mathrm{P}\left(\text{Y=B} \mid \text{X=A and Z=C}\right) \\ &= \mathrm{P}\left(X = A \,|\, Z = C\right) \mathrm{P}\left(Y = B \,|\, Z = C\right) \end{aligned}$$

Conditional independence

If votes are conditionally independent given affiliation

$$\mathrm{P}(\text{V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N} \mid \text{Aff} = \text{R})$$
$$= \mathrm{P}(\text{V1} = \text{Y} \mid \text{Aff} = \text{R})\, \mathrm{P}(\text{V2} = \text{Y} \mid \text{Aff} = \text{R}) \ldots \mathrm{P}(\text{V16} = \text{N} \mid \text{Aff} = \text{R})$$

$$\mathrm{P}(\text{V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N} \mid \text{Aff} = \text{D})$$
$$= \mathrm{P}(\text{V1} = \text{Y} \mid \text{Aff} = \text{D})\, \mathrm{P}(\text{V2} = \text{Y} \mid \text{Aff} = \text{D}) \ldots \mathrm{P}(\text{V16} = \text{N} \mid \text{Aff} = \text{D})$$

Are votes conditionally independent?

$$\mathrm{P}\left(\text{Vote 4= Yes} \mid \text{Aff} = \text{R}\right) = 0.952$$

$$\mathrm{P}\left(\text{Vote 11 = Yes} \mid \text{Aff} = \text{R}\right) = 0.871$$

$$\mathrm{P}\left(\text{Vote 4= Yes and Vote 11 = Yes} \mid \text{Aff} = \text{R}\right) = 0.851$$

Are votes conditionally independent?

$$\mathrm{P}\left(\text{Vote 4= Yes} \mid \text{Aff} = \text{R}\right) = 0.952$$

$$\mathrm{P}\left(\text{Vote 11} = \text{Yes}\mid \text{Aff} = \text{R}\right) = 0.871$$

$$\mathrm{P}\left(\text{Vote 4= Yes and Vote 11} = \text{Yes}\mid \text{Aff} = \text{R}\right) = 0.851$$

$$\mathrm{P}\left(\text{Vote 4= Yes} \mid \text{Aff} = \text{R}\right)\mathrm{P}\left(\text{Vote 11} = \text{Yes} \mid \text{Aff} = \text{R}\right) =$$

Are votes conditionally independent?

$$\mathrm{P}\left(\text{Vote 4= Yes} \mid \text{Aff} = \text{R}\right) = 0.952$$

$$\mathrm{P}\left(\text{Vote 11 = Yes} \mid \text{Aff} = \text{R}\right) = 0.871$$

$$\mathrm{P}\left(\text{Vote 4= Yes and Vote 11 = Yes} \mid \text{Aff} = \text{R}\right) = 0.851$$

$$\mathrm{P}\left(\text{Vote 4= Yes} \mid \text{Aff} = \text{R}\right) \mathrm{P}\left(\text{Vote 11 = Yes} \mid \text{Aff} = \text{R}\right) = 0.829$$

Are votes conditionally independent?

$$\mathrm{P}\,(\text{Vote 4= Yes} \mid \text{Aff = D}) = 0.216$$

$$\mathrm{P}\,(\text{Vote 11 = Yes} \mid \text{Aff = D}) = 0.145$$

$$\mathrm{P}\,(\text{Vote 4= Yes and Vote 11 = Yes} \mid \text{Aff = D}) = 0.075$$

Are votes conditionally independent?

$$\mathrm{P}\left(\text{Vote 4= Yes} \mid \text{Aff} = \text{D}\right) = 0.216$$

$$\mathrm{P}\left(\text{Vote 11 = Yes} \mid \text{Aff} = \text{D}\right) = 0.145$$

$$\mathrm{P}\left(\text{Vote 4= Yes and Vote 11 = Yes} \mid \text{Aff} = \text{D}\right) = 0.075$$

$$\mathrm{P}\left(\text{Vote 4= Yes} \mid \text{Aff} = \text{D}\right) \mathrm{P}\left(\text{Vote 11 = Yes} \mid \text{Aff} = \text{D}\right) =$$

Are votes conditionally independent?

$$\mathrm{P}\left(\text{Vote 4= Yes} \mid \text{Aff = D}\right) = 0.216$$

$$\mathrm{P}\left(\text{Vote 11 = Yes}\mid \text{Aff = D}\right) = 0.145$$

$$\mathrm{P}\left(\text{Vote 4= Yes and Vote 11 = Yes}\mid \text{Aff = D}\right) = 0.075$$

$$\mathrm{P}\left(\text{Vote 4= Yes} \mid \text{Aff = D}\right)\mathrm{P}\left(\text{Vote 11 = Yes} \mid \text{Aff = D}\right) = 0.031$$

Conditional probability of YES given affiliation

	V1	V2	V3	V4	V5	V6	V7	V8
R	0.19	0.50	0.14	0.99	0.95	0.90	0.24	0.15
D	0.61	0.50	0.89	0.05	0.22	0.47	0.78	0.83

	V9	V10	V11	V12	V13	V14	V15	V16
R	0.11	0.55	0.14	0.87	0.86	0.98	0.09	0.66
D	0.76	0.47	0.51	0.15	0.29	0.35	0.64	0.94

Predicting affiliation

By the chain rule

$$\mathrm{P}\left(\text{Aff} = \text{R} \mid \text{V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N}\right)$$
$$= \frac{\mathrm{P}\left(\text{Aff} = \text{R and V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N}\right)}{\mathrm{P}\left(\text{V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N}\right)}$$

Approximation

$$
\begin{aligned}
&\mathrm{P}(\text{Aff} = \text{R and V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N}) \\
&\quad = \mathrm{P}(\text{Aff} = \text{R})\,\mathrm{P}(\text{V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N} \mid \text{Aff} = \text{R}) \\
&\quad \approx \mathrm{P}(\text{Aff} = \text{R})\,\mathrm{P}(\text{V1} = \text{Y} \mid \text{Aff} = \text{R})\,\mathrm{P}(\text{V2} = \text{Y} \mid \text{Aff} = \text{R}) \ldots \mathrm{P}(\text{V16} = \text{N} \mid \text{Aff} = \text{R})
\end{aligned}
$$

Approximation

$$\mathrm{P}(\text{Aff} = \text{R and V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N})$$
$$= \mathrm{P}(\text{Aff} = \text{R})\, \mathrm{P}(\text{V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N} \mid \text{Aff} = \text{R})$$
$$\approx \mathrm{P}(\text{Aff} = \text{R})\, \mathrm{P}(\text{V1} = \text{Y} \mid \text{Aff} = \text{R})\, \mathrm{P}(\text{V2} = \text{Y} \mid \text{Aff} = \text{R}) \ldots \mathrm{P}(\text{V16} = \text{N} \mid \text{Aff} = \text{R})$$

What about

$$\mathrm{P}(\text{V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N}) \ ?$$

Law of total probability

If X equals A or B

$$\mathrm{P}\left(Y = C\right) = \mathrm{P}\left(Y = C \text{ and } X = A\right) + \mathrm{P}\left(Y = C \text{ and } X = B\right)$$

Law of total probability

If X equals A or B

$$\mathrm{P}(Y = C) = \mathrm{P}(Y = C \text{ and } X = A) + \mathrm{P}(Y = C \text{ and } X = B)$$

Proof:

$$\mathrm{P}(Y = C \text{ and } X = A) + \mathrm{P}(Y = C \text{ and } X = B)$$

$$=$$

Law of total probability

If X equals A or B

$$\mathrm{P}\left(Y = C\right) = \mathrm{P}\left(Y = C \text{ and } X = A\right) + \mathrm{P}\left(Y = C \text{ and } X = B\right)$$

Proof:

$$\begin{aligned}&\mathrm{P}\left(Y = C \text{ and } X = A\right) + \mathrm{P}\left(Y = C \text{ and } X = B\right)\\&= \frac{\#Y = C \text{ and } X = A}{\text{Total}} + \frac{\#Y = C \text{ and } X = B}{\text{Total}}\end{aligned}$$

Law of total probability

If X equals A or B

$$\mathrm{P}(Y = C) = \mathrm{P}(Y = C \text{ and } X = A) + \mathrm{P}(Y = C \text{ and } X = B)$$

Proof:

$$\begin{aligned}
&\mathrm{P}(Y = C \text{ and } X = A) + \mathrm{P}(Y = C \text{ and } X = B) \\
&\quad = \frac{\#Y = C \text{ and } X = A}{\text{Total}} + \frac{\#Y = C \text{ and } X = B}{\text{Total}} \\
&\quad = \frac{\#Y = C}{\text{Total}}
\end{aligned}$$

Law of total probability

If X equals A or B

$$\mathrm{P}\left(Y = C\right) = \mathrm{P}\left(Y = C \text{ and } X = A\right) + \mathrm{P}\left(Y = C \text{ and } X = B\right)$$

Proof:

$$\begin{aligned}
&\mathrm{P}\left(Y = C \text{ and } X = A\right) + \mathrm{P}\left(Y = C \text{ and } X = B\right) \\
&\quad = \frac{\#Y = C \text{ and } X = A}{\text{Total}} + \frac{\#Y = C \text{ and } X = B}{\text{Total}} \\
&\quad = \frac{\#Y = C}{\text{Total}} \\
&\quad = \mathrm{P}\left(Y = C\right)
\end{aligned}$$

Law of total probability

$$
\begin{aligned}
&\mathrm{P}\left(\text{V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N}\right)\\
&\quad = \mathrm{P}\left(\text{Aff} = \text{R and V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N}\right)\\
&\qquad + \mathrm{P}\left(\text{Aff} = \text{D and V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N}\right)
\end{aligned}
$$

Law of total probability

$$\begin{aligned}
&\mathrm{P}\left(\text{V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N}\right) \\
&\quad = \mathrm{P}\left(\text{Aff} = \text{R and V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N}\right) \\
&\qquad + \mathrm{P}\left(\text{Aff} = \text{D and V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N}\right)
\end{aligned}$$

$$\begin{aligned}
&\mathrm{P}\left(\text{Aff} = \text{R and V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N}\right) \\
&\quad {\color{red}\approx} \mathrm{P}\left(\text{Aff} = \text{R}\right) \mathrm{P}\left(\text{V1} = \text{Y} \mid \text{Aff} = \text{R}\right) \mathrm{P}\left(\text{V2} = \text{Y} \mid \text{Aff} = \text{R}\right) \ldots \mathrm{P}\left(\text{V16} = \text{N} \mid \text{Aff} = \text{R}\right)
\end{aligned}$$

$$\begin{aligned}
&\mathrm{P}\left(\text{Aff} = \text{D and V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N}\right) \\
&\quad {\color{red}\approx} \mathrm{P}\left(\text{Aff} = \text{D}\right) \mathrm{P}\left(\text{V1} = \text{Y} \mid \text{Aff} = \text{D}\right) \mathrm{P}\left(\text{V2} = \text{Y} \mid \text{Aff} = \text{D}\right) \ldots \mathrm{P}\left(\text{V16} = \text{N} \mid \text{Aff} = \text{D}\right)
\end{aligned}$$

Naive Bayes

$$\mathrm{P}\left(\mathsf{Aff} = \mathsf{R} \mid \mathsf{V1} = \mathsf{Y} \text{ and } \mathsf{V2} = \mathsf{Y} \text{ and } \ldots \mathsf{V16} = \mathsf{N}\right)$$
$$= \frac{\mathrm{P}\left(\mathsf{Aff} = \mathsf{R} \text{ and } \mathsf{V1} = \mathsf{Y} \text{ and } \mathsf{V2} = \mathsf{Y} \text{ and } \ldots \mathsf{V16} = \mathsf{N}\right)}{\mathrm{P}\left(\mathsf{V1} = \mathsf{Y} \text{ and } \mathsf{V2} = \mathsf{Y} \text{ and } \ldots \mathsf{V16} = \mathsf{N}\right)}$$

Naive Bayes

$$\mathrm{P}\left(\text{Aff} = \text{R} \mid \text{V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N}\right)$$
$$= \frac{\mathrm{P}\left(\text{Aff} = \text{R and V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N}\right)}{\mathrm{P}\left(\text{V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N}\right)}$$

$$\mathrm{P}\left(\text{Aff} = \text{R and V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N}\right)$$
$$\approx \mathrm{P}\left(\text{Aff} = \text{R}\right) \mathrm{P}\left(\text{V1} = \text{Y} \mid \text{Aff} = \text{R}\right) \mathrm{P}\left(\text{V2} = \text{Y} \mid \text{Aff} = \text{R}\right) \ldots \mathrm{P}\left(\text{V16} = \text{N} \mid \text{Aff} = \text{R}\right)$$

Naive Bayes

$$\mathrm{P}(\text{Aff} = \text{R} \mid \text{V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N})$$
$$= \frac{\mathrm{P}(\text{Aff} = \text{R and V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N})}{\mathrm{P}(\text{V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N})}$$

$$\mathrm{P}(\text{Aff} = \text{R and V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N})$$
$$\approx \mathrm{P}(\text{Aff} = \text{R})\,\mathrm{P}(\text{V1} = \text{Y} \mid \text{Aff} = \text{R})\,\mathrm{P}(\text{V2} = \text{Y} \mid \text{Aff} = \text{R}) \ldots \mathrm{P}(\text{V16} = \text{N} \mid \text{Aff} = \text{R})$$

$$\mathrm{P}(\text{Aff} = \text{R and V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N})$$
$$= \mathrm{P}(\text{Aff} = \text{R and V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N})$$
$$+ \mathrm{P}(\text{Aff} = \text{D and V1} = \text{Y and V2} = \text{Y and} \ldots \text{V16} = \text{N})$$

Naive Bayes

$$\mathrm{P}\left(\textsf{Aff} = \textsf{R} \mid \textsf{V1} = \textsf{Y and V2} = \textsf{Y and} \ldots \textsf{V16} = \textsf{N}\right) = \frac{\mathrm{P}\left(\textsf{Aff} = \textsf{R and V1} = \textsf{Y and V2} = \textsf{Y and} \ldots \textsf{V16} = \textsf{N}\right)}{\mathrm{P}\left(\textsf{V1} = \textsf{Y and V2} = \textsf{Y and} \ldots \textsf{V16} = \textsf{N}\right)}$$

$$\begin{aligned}&\mathrm{P}\left(\textsf{Aff} = \textsf{R and V1} = \textsf{Y and V2} = \textsf{Y and} \ldots \textsf{V16} = \textsf{N}\right)\\ &\quad \approx \mathrm{P}\left(\textsf{Aff} = \textsf{R}\right)\mathrm{P}\left(\textsf{V1} = \textsf{Y} \mid \textsf{Aff} = \textsf{R}\right)\mathrm{P}\left(\textsf{V2} = \textsf{Y} \mid \textsf{Aff} = \textsf{R}\right)\ldots\mathrm{P}\left(\textsf{V16} = \textsf{N} \mid \textsf{Aff} = \textsf{R}\right)\end{aligned}$$

$$\begin{aligned}&\mathrm{P}\left(\textsf{Aff} = \textsf{R and V1} = \textsf{Y and V2} = \textsf{Y and} \ldots \textsf{V16} = \textsf{N}\right)\\ &\quad = \mathrm{P}\left(\textsf{Aff} = \textsf{R and V1} = \textsf{Y and V2} = \textsf{Y and} \ldots \textsf{V16} = \textsf{N}\right)\\ &\qquad + \mathrm{P}\left(\textsf{Aff} = \textsf{D and V1} = \textsf{Y and V2} = \textsf{Y and} \ldots \textsf{V16} = \textsf{N}\right)\end{aligned}$$

$$\begin{aligned}&\mathrm{P}\left(\textsf{Aff} = \textsf{D and V1} = \textsf{Y and V2} = \textsf{Y and} \ldots \textsf{V16} = \textsf{N}\right)\\ &\quad \approx \mathrm{P}\left(\textsf{Aff} = \textsf{D}\right)\mathrm{P}\left(\textsf{V1} = \textsf{Y} \mid \textsf{Aff} = \textsf{D}\right)\mathrm{P}\left(\textsf{V2} = \textsf{Y} \mid \textsf{Aff} = \textsf{D}\right)\ldots\mathrm{P}\left(\textsf{V16} = \textsf{N} \mid \textsf{Aff} = \textsf{D}\right)\end{aligned}$$

Results

- We approximate probabilities from 425 politicians
- We predict affiliation of 10 other politicians
- For 3, probability of republican ≈ 0 (truth:)
- For 7, probability of republican ≈ 1 (truth:)

Results

- We approximate probabilities from 425 politicians
- We predict affiliation of 10 other politicians
- For 3, probability of republican ≈ 0 (truth: all are democrats)
- For 7, probability of republican ≈ 1 (truth:)

Results

- We approximate probabilities from 425 politicians
- We predict affiliation of 10 other politicians
- For 3, probability of republican ≈ 0 (truth: all are democrats)
- For 7, probability of republican ≈ 1 (truth: one is a democrat!)

Error

	V1	V2	V3	V4	V5	V6	V7	V8
R	0.19	0.50	0.14	0.99	0.95	0.90	0.24	0.15
D	0.61	0.50	0.89	0.05	0.22	0.47	0.78	0.83
E	Y	Y	–	Y	Y	Y	N	N

	V9	V10	V11	V12	V13	V14	V15	V16
R	0.11	0.55	0.14	0.87	0.86	0.98	0.09	0.66
D	0.76	0.47	0.51	0.15	0.29	0.35	0.64	0.94
E	Y	N	Y	–	Y	Y	N	N