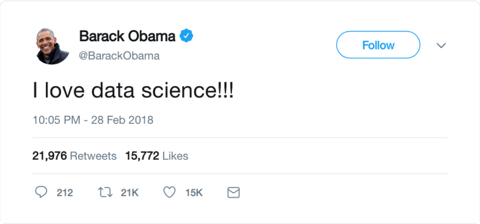


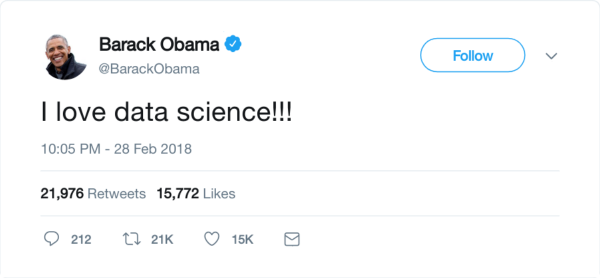
DSC 40A

Lecture 17

Naïve Bayes, pt II

Sentiment Analysis

- Goal: given a tweet, determine if it is **positive**, **negative**, or **neutral**.



Classification

We have collected data:

Sentiment	"Love"	"Hate"	"Good"
positive	yes	no	yes
positive	no	no	yes
positive	yes	yes	no
negative	yes	yes	no
negative	no	no	yes
neutral	no	no	no
neutral	no	no	no

We want to classify a new tweet as **positive**, **negative**, or **neutral**:

Conditional Probabilities for Classification

- We will calculate:

$$P(\textbf{positive} \mid \text{love=no \& hate=yes \& good=yes})$$
$$P(\textbf{negative} \mid \text{love=no \& hate=yes \& good=yes})$$
$$P(\textbf{neutral} \mid \text{love=no \& hate=yes \& good=yes})$$

- We will classify the tweet as whichever sentiment whose conditional probability is largest.

Approximating Probabilities with a Sample

P(**positive** | love=no & hate=yes & good=yes)

$$= \frac{\text{\# of \textbf{positive} tweets with hate and good, but not love}}{\text{\# of tweets with hate and good, but not love}}$$

$$\approx \frac{\text{\# of \textbf{positive} tweets in sample with hate and good, but not love}}{\text{\# of tweets in sample with hate and good, but not love}}$$

The Curse of Dimensionality

P(**positive** | love=no & hate=yes & good=yes)

$$\approx \frac{\text{\# of \textbf{positive} tweets in sample with hate and good, but not love}}{\text{\# of tweets in sample with hate and good, but not love}}$$

=

Sentiment	"Love"	"Hate"	"Good"
positive	yes	no	yes
positive	no	no	yes
positive	yes	yes	no
negative	yes	yes	no
negative	no	no	yes
neutral	no	no	no
neutral	no	no	no

Naïve Bayes

To approximate $P(\textbf{positive} \mid \text{love=no \& hate=yes \& good=yes})$:

- Start with Bayes' Theorem:

$$P(\textbf{positive} \mid \text{love=no \& hate=yes \& good=yes})$$

$$= \frac{P(\text{love=no \& hate=yes \& good=yes} \mid \textbf{positive}) \cdot P(\textbf{positive})}{P(\text{love=no \& hate=yes \& good=yes})}$$

- Then **assume** conditional independence of features given class:

$$P(\text{love=no \& hate=yes \& good=yes} \mid \textbf{positive})$$

$$= P(\text{love=no} \mid \textbf{positive}) \cdot P(\text{hate=yes} \mid \textbf{positive}) \cdot P(\text{good=yes} \mid \textbf{positive})$$

Naïve Bayes

$$P(\textbf{positive} \mid \text{love=no \& hate=yes \& good=yes})$$

$$= \frac{P(\text{love=no \& hate=yes \& good=yes} \mid \textbf{positive}) \cdot P(\textbf{positive})}{P(\text{love=no \& hate=yes \& good=yes})}$$

$$= \frac{P(\text{love=no} \mid \textbf{positive}) \cdot P(\text{hate=yes} \mid \textbf{positive}) \cdot P(\text{good=yes} \mid \textbf{positive}) \cdot P(\textbf{positive})}{P(\text{love=no \& hate=yes \& good=yes})}$$

- We *are* able to estimate $P(\text{love=no} \mid \textbf{positive})$, $P(\text{hate=yes} \mid \textbf{positive})$, and $P(\text{good=yes} \mid \textbf{positive})$ from the data.

Naïve Bayes

We want to find the biggest out of:

$$P(\textbf{positive} \mid \text{love=no \& hate=yes \& good=yes})$$

$$= \frac{P(\text{love=no} \mid \textbf{positive}) \cdot P(\text{hate=yes} \mid \textbf{positive}) \cdot P(\text{good=yes} \mid \textbf{positive}) \cdot P(\textbf{positive})}{P(\text{love=no \& hate=yes \& good=yes})}$$

$$P(\textbf{negative} \mid \text{love=no \& hate=yes \& good=yes})$$

$$= \frac{P(\text{love=no} \mid \textbf{negative}) \cdot P(\text{hate=yes} \mid \textbf{negative}) \cdot P(\text{good=yes} \mid \textbf{negative}) \cdot P(\textbf{negative})}{P(\text{love=no \& hate=yes \& good=yes})}$$

$$P(\textbf{neutral} \mid \text{love=no \& hate=yes \& good=yes})$$

$$= \frac{P(\text{love=no} \mid \textbf{neutral}) \cdot P(\text{hate=yes} \mid \textbf{neutral}) \cdot P(\text{good=yes} \mid \textbf{neutral}) \cdot P(\textbf{neutral})}{P(\text{love=no \& hate=yes \& good=yes})}$$

Naïve Bayes

Since they all have the same denominator, we can just pick that with the largest numerator:

$$P(\text{love=no} \mid \textbf{positive}) \cdot P(\text{hate=yes} \mid \textbf{positive}) \cdot P(\text{good=yes} \mid \textbf{positive}) \cdot P(\textbf{positive})$$

$$P(\text{love=no} \mid \textbf{negative}) \cdot P(\text{hate=yes} \mid \textbf{negative}) \cdot P(\text{good=yes} \mid \textbf{negative}) \cdot P(\textbf{negative})$$

$$P(\text{love=no} \mid \textbf{neutral}) \cdot P(\text{hate=yes} \mid \textbf{neutral}) \cdot P(\text{good=yes} \mid \textbf{neutral}) \cdot P(\textbf{neutral})$$

This is **Naïve Bayes** classification.

Running Naïve Bayes

$P(\text{love=no} \mid \textbf{positive}) \cdot P(\text{hate=yes} \mid \textbf{positive}) \cdot P(\text{good=yes} \mid \textbf{positive}) \cdot P(\textbf{positive})$

Sentiment	"Love"	"Hate"	"Good"
positive	yes	no	yes
positive	no	no	yes
positive	yes	yes	no
negative	yes	yes	no
negative	no	no	yes
neutral	no	no	no
neutral	no	no	no

Running Naïve Bayes

$P(\text{love=no} \mid \textbf{negative}) \cdot P(\text{hate=yes} \mid \textbf{negative}) \cdot P(\text{good=yes} \mid \textbf{negative}) \cdot P(\textbf{negative})$

Sentiment	"Love"	"Hate"	"Good"
positive	yes	no	yes
positive	no	no	yes
positive	yes	yes	no
negative	yes	yes	no
negative	no	no	yes
neutral	no	no	no
neutral	no	no	no

Running Naïve Bayes

$P(\text{love=no} \mid \textbf{neutral}) \cdot P(\text{hate=yes} \mid \textbf{neutral}) \cdot P(\text{good=yes} \mid \textbf{neutral}) \cdot P(\textbf{neutral})$

Sentiment	“Love”	“Hate”	“Good”
positive	yes	no	yes
positive	no	no	yes
positive	yes	yes	no
negative	yes	yes	no
negative	no	no	yes
neutral	no	no	no
neutral	no	no	no

The Classification

We have:

$$P(\text{love=no} \mid \textbf{positive}) \cdot P(\text{hate=yes} \mid \textbf{positive}) \cdot P(\text{good=yes} \mid \textbf{positive}) \cdot P(\textbf{positive})$$

$$\approx$$

$$P(\text{love=no} \mid \textbf{negative}) \cdot P(\text{hate=yes} \mid \textbf{negative}) \cdot P(\text{good=yes} \mid \textbf{negative}) \cdot P(\textbf{negative})$$

$$\approx$$

$$P(\text{love=no} \mid \textbf{neutral}) \cdot P(\text{hate=yes} \mid \textbf{neutral}) \cdot P(\text{good=yes} \mid \textbf{neutral}) \cdot P(\textbf{neutral})$$

$$\approx$$

So we classify the tweet as: **positive**, **negative**, **neutral**.

Example

Suppose that today's humidity is > 50%, the temperature is hot, and the air pressure is low. Use Naïve Bayes to predict whether tomorrow will be rainy, cloudy, or sunny.

Next Day's Weather	Humidity	Temperature	Air Pressure
Rainy	> 50%	Cool	Low
Rainy	> 50%	Hot	Low
Rainy	> 50%	Cool	Low
Rainy	25%-50%	Hot	High
Rainy	25%-50%	Hot	Low
Rainy	25%-50%	Cool	Low
Rainy	25%-50%	Cool	Low
Rainy	< 25%	Cool	Low
Rainy	< 25%	Hot	Low
Rainy	< 25%	Hot	High
Cloudy	> 50%	Cool	Low
Cloudy	> 50%	Cool	Low
Cloudy	25%-50%	Hot	High
Cloudy	< 25%	Cool	High
Cloudy	< 25%	Cool	Low
Sunny	> 50%	Cool	Low
Sunny	> 50%	Hot	High
Sunny	> 50%	Cool	High
Sunny	25%-50%	Hot	High
Sunny	< 25%	Hot	High