Naive Bayes, Laplace Smoothing, and Perceptron Learning

Review Examples

Naive Bayes Model

- Want to model the joint distribution of our features **X** and label Y:

$$P(Y, \mathbf{X}) = P(Y = y, X_1 = x_1, X_2 = x_2, \ldots, X_d, = x_d)$$

- Naive Bayes independence assumption
  - All features are independent of each other, given Y

$$P(Y, \mathbf{X}) = P(Y = y) \prod_{i=1}^{d} P(X_i = x_i | Y = y)$$

  - Choose the most likely Y $\quad y^* = \arg\max_y P(Y = y) \prod_{i=1}^{d} P(X_i = x_i | Y = y$

- Maximum Likelihood estimates for model parameters

$$P(X_i = x_i) = \frac{count(X_i = x_i)}{N}$$

  - where N=#examples in data

Laplace Smoothing

- What if you've never seen feature before?
  - Joint probability becomes 0
  - Just because an event has not happened before, does not mean that it won't ever happen.
    - Use Laplace smoothing for Max Likelihood estimates. "Pretend you've seen each variable k extra times"

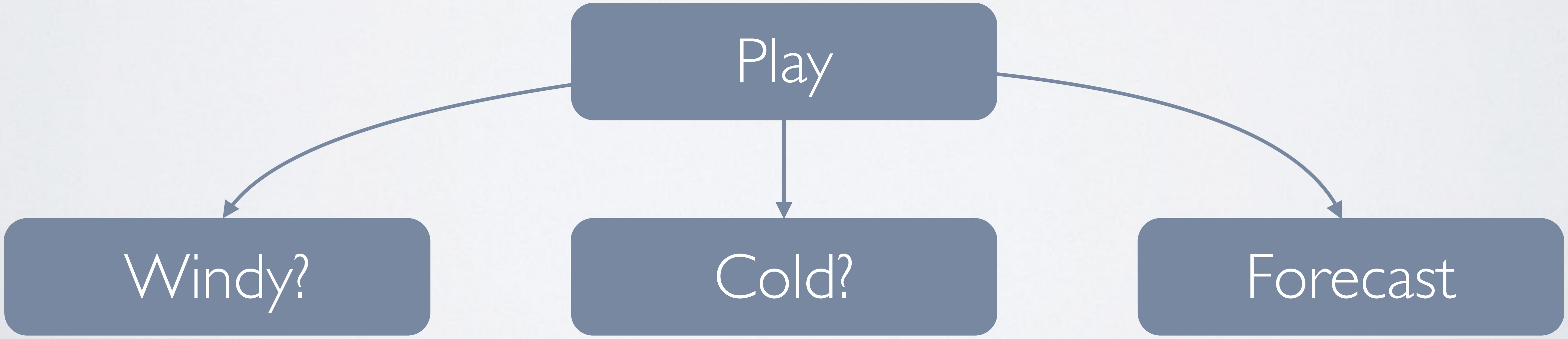
$$P_{Lap,k}(X_i = x_i) = \frac{count(X_i = x_i) + k}{N + k|X_i|} \qquad P_{Lap,k}(X_i = x_i | Y = y) = \frac{count(X_i = x_i, Y = y) + k}{count(Y = y) + k|X_i|}$$

  - where $|X_i|$ = # values X_i can take on (i.e. 2 for binary, and k is the Laplace smoothing factor

NB + Laplace Example

Day	Windy?	Cold?	Forecast	Play
1	T	T	Sunny	T
2	T	T	Cloudy	F
3	T	F	Sunny	T
4	F	T	Rainy	F
5	F	F	Rainy	F
6	T	F	Cloudy	T
7	F	T	Rainy	F
8	F	T	Sunny	T

Weather data and decision to play tennis from 8 different days



NB Estimates w/ Laplace Smoothing

Play	P(Play)
T	$\frac{4+1}{8+2} = 0.5$
F	$\frac{4+1}{8+2} = 0.5$

Windy?	Play	P(Windy? | Play)
T	T	$\frac{3+1}{4+2} = 0.67$
T	F	$\frac{1+1}{4+2} = 0.33$
F	T	$\frac{1+1}{4+2} = 0.33$
F	F	$\frac{3+1}{4+2} = 0.67$

Cold?	Play	P(Cold? | Play)
T	T	$\frac{2+1}{4+2} = 0.5$
T	F	$\frac{3+1}{4+2} = 0.67$
F	T	$\frac{2+1}{4+2} = 0.5$
F	F	$\frac{1+1}{4+2} = 0.33$

Forecast	Play	P(Forecast | Play)
Sunny	T	$\frac{3+1}{4+3} = 0.57$
Sunny	F	$\frac{0+1}{4+3} = 0.14$
Cloudy	T	$\frac{1+1}{4+3} = 0.29$
Cloudy	F	$\frac{1+1}{4+3} = 0.29$
Rainy	T	$\frac{0+1}{4+3} = 0.14$
Rainy	F	$\frac{3+1}{4+3} = 0.57$

NB + Laplace Example

Using Laplace smoothed (k=1) NB estimates,
what is the probability that they **will Play**, with
Windy = T, **Cold = F,** and **Forecast = Rainy** ?

P(Play=T)*P(Windy=T | Play=T) * P(Cold = F | Play=T) * P(Forecast=Rainy | Play=T)

= (0.5)(0.67)(0.5) (0.14)
= 0.02

Perceptron Learning

- Prediction function: $h(\mathbf{x}) = \texttt{sign}(\mathbf{w} \cdot \mathbf{x})$

$$= \begin{cases} +1 & \mathbf{w} \cdot \mathbf{x} \geq 0 \\ -1 & \text{otherwise} \end{cases}$$

- Weight Updates
  - Given:
    - Current weight vector at time t: $\mathbf{w}_t$
    - Example at time t+1: ($\mathbf{x}$, y*)
  - Update
    - If prediction correct (h($\mathbf{x}$) = y*), $\mathbf{w}_{t+1} = \mathbf{w}_t$
    - Else, $\mathbf{w}_{t+1} = \mathbf{w}_t + (\alpha y^*)\mathbf{x}$
      - α is learning rate (or step size)

Perceptron Example

- Given initial weight vector $\mathbf{w}_1 = \langle 1, 0, 2 \rangle$
  What is the weight vector after seeing both of the following two examples (assume learning rate of 1)?

$$(\langle 1, 4, -1 \rangle, +1)$$

$$(\langle -1, 1, 0 \rangle, -1)$$

Perceptron Example

- Example 1: $(\langle 1, 4, -1\rangle, +1)$ $\qquad \mathbf{w}_1 = \langle 1, 0, 2\rangle$

$$\begin{aligned} h(\mathbf{x}_1) &= \texttt{sign}(\mathbf{w}_1 \cdot \mathbf{x}_1) \\ &= \texttt{sign}\left(1(1) + 0(4) + 2(-1)\right) \\ &= \texttt{sign}(-1) \\ &= -1 \end{aligned}$$

- Incorrect $h(\mathbf{x}_1) = -1 \neq y_1^*$
- Update weights: $\mathbf{w}_2 = \mathbf{w}_1 + \alpha y_1^*(\mathbf{x}_1)$

$$\begin{aligned} &= \langle 1, 0, 2\rangle + 1(1)\langle 1, 4, -1\rangle \\ &= \langle 2, 4, 1\rangle \end{aligned}$$

Perceptron Example

- Example 1: $(\langle -1, 1, 0\rangle, -1)$

$\mathbf{w}_2 = \langle 2, 4, 1\rangle$

$$\begin{aligned} h(\mathbf{x}_2) &= \texttt{sign}(\mathbf{w}_2 \cdot \mathbf{x}_2) \\ &= \texttt{sign}\left(2(-1) + 4(1) + 1(0)\right) \\ &= \texttt{sign}(2) \\ &= +1 \end{aligned}$$

- Incorrect $h(\mathbf{x}_2) = +1 \neq y_2^*$
- Update weights: $$\begin{aligned} \mathbf{w}_3 &= \mathbf{w}_2 + \alpha y_2^*(\mathbf{x}_2) \\ &= \langle 2, 4, 1\rangle + 1(-1)\langle -1, 1, 0\rangle \\ &= \langle 3, 3, 1\rangle \end{aligned}$$