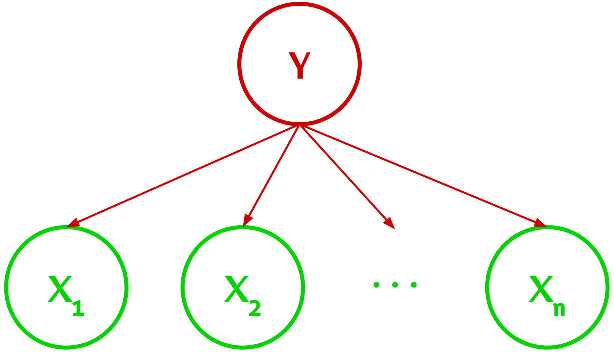
Exam Prep Discussion 11

Maximum Likelihood Estimation

- Setup: n observations $X_1, \ldots, X_n$ drawn from a distribution parameterized by θ
  - We know that θ "determines" the X_i's
  - Now that we've seen the X_i's, what is the most likely value of θ?

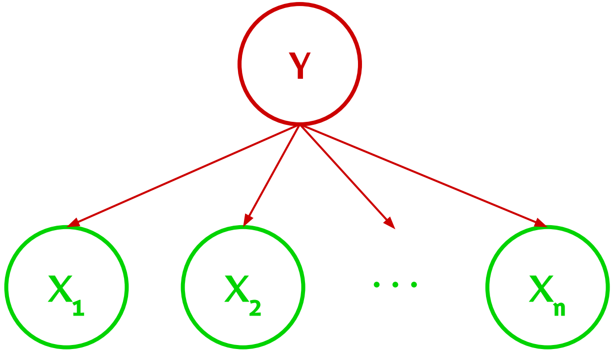
Naive Bayes



Naive Bayes

- Method to predict the probability of observing a label Y given features X_1, ..., X_n:

$$P(Y = y | X_1 = x_1, \dots, X_n = x_n)$$



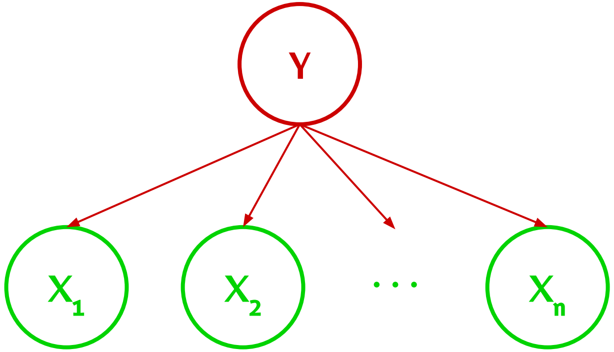
Naive Bayes

- Method to predict the probability of observing a label Y given features $X_1, ..., X_n$:

$$P(Y = y | X_1 = x_1, \ldots, X_n = x_n)$$

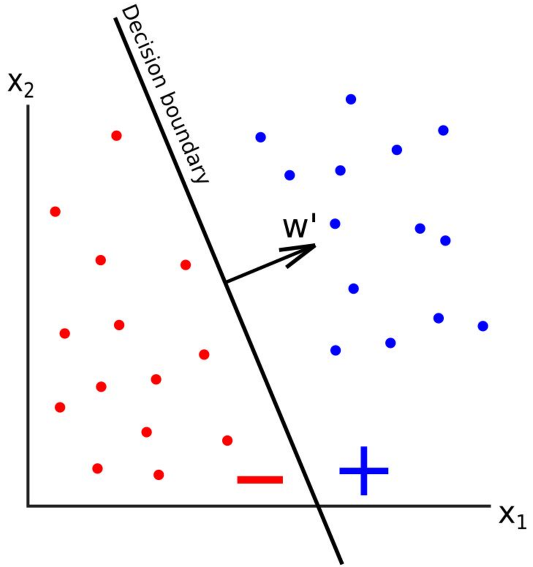
- Prediction formula:

$$\text{Pred}(X) = \text{argmax}_y P(Y = y | X_1 = x_1, \ldots, X_n = x_n)$$



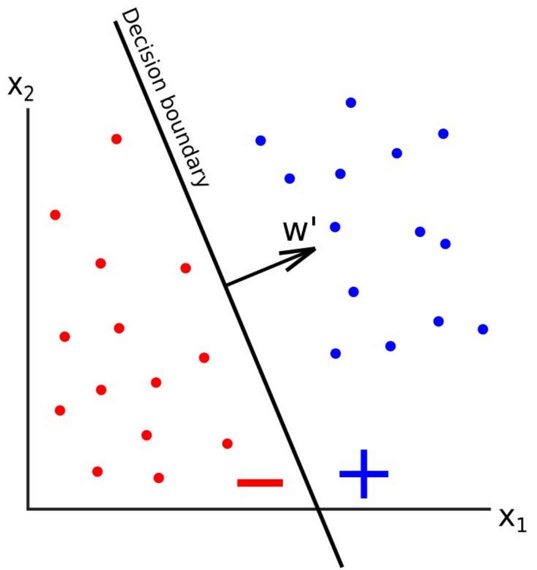
Perceptron

- Assume data $\vec{x} = [x_1, \ldots, x_d]^T$ and has a label y^*
- Machine learning algorithm that generates a linear decision boundary in d dimensions
  - Uses weight vector $\vec{w} = [w_1, \ldots, w_d]^T$
- Goal: $\vec{w}^T \vec{x}$ correctly determines the sign of y^*



Perceptron

- Assume data $\vec{x} = [x_1, \ldots, x_d]^T$ and has a label y^*
- Machine learning algorithm that generates a linear decision boundary in d dimensions
  - Uses weight vector $\vec{w} = [w_1, \ldots, w_d]^T$
- Goal: $\vec{w}^T\vec{x}$ correctly determines the sign of y^*



Algorithm.

1) Start with $\vec{w} = 0$
2) While not all data points are correctly classified,
   a) Predict label of point using current weight vector:

$$y_{\text{pred}} = \begin{cases} +1 & w^T x \geq 0 \\ -1 & w^T x < 0 \end{cases}$$

   b) If prediction is *incorrect*, update weight vector:

$$\vec{w} \leftarrow \vec{w} + \alpha y^* \vec{x}$$

x: [2, 1], with label y* = -1. α = 1
w = [1, 0]
$w^Tx = 2 + 0 = 2 \Rightarrow y_{pred} = +1$
w' = [1, 0] + (-1)[2, 1] = [-1, -1]
Now, $w'^Tx = -2 - 1 = -3 \Rightarrow y_{pred} = -1 = y^*$