CS 4824/ECE 4424: GNB Decision Surface

Acknowledgement:
Many of these slides are derived from Tom Mitchell, Pascal Poupart, Pieter Abbeel, Eric Eaton, Carlos Guestrin, William Cohen, and Andrew Moore.
Gaussian Naïve Bayes - decision surface

- Assume $Y=\text{PlayBasketball}$ (boolean) $X_1=\text{Height}$ $X_2=\text{Age}$
  - $Y^{\text{New}} \leftarrow \arg \max_{y_k} P(Y \mid y_k) \prod_i P(X_i^{\text{New}} \mid Y = y_k)$; assume $P(Y=1) = 0.5$
What is the minimum possible error?

- Best case:
  - Conditional independence assumption is satisfied
  - We can perfectly estimate $P(Y)$, $P(X|Y)$ (i.e. infinite training data)

$$
Pr.(err) = \Pr.(\text{predicting } Y=0, \text{ but } Y=1) + \Pr.(\text{predicting } Y=1, \text{ but } Y=0)
$$

$$
= \int P(X_1|Y=1) P(Y=1) \, dt + \int P(X_1|Y=0) P(Y=0) \, dt
$$
But...

- Naïve Bayes allows estimating $P(Y|X)$ by learning $P(Y)$ and $P(X|Y)$
- Why not learn $P(Y|X)$ directly?