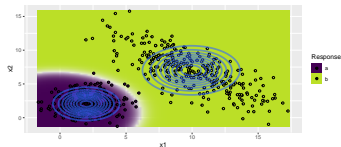
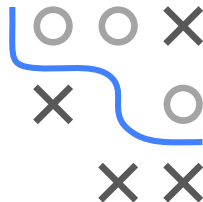


## Classification: Naive Bayes



- Understand the idea of Naive Bayes
- Understand in which sense Naive Bayes is a special QDA model

- Understand the idea of Naive Bayes
- Understand in which sense Naive Bayes is a special QDA model





## NB: CATEGORICAL FEATURES

We use a categorical distribution for  $p(x_j|y = k)$  and estimate the probabilities  $p_{kjm}$  that, in class  $k$ , our  $j$ -th feature has value  $m$ ,  $x_j = m$ , simply by counting the frequencies.

$$p(x_j|y = k) = \prod_m p_{kjm}^{[x_j=m]}$$

Because of the simple conditional independence structure it is also very easy to deal with mixed numerical / categorical feature spaces.



# LAPLACE SMOOTHING

If a given class and feature value never occur together in the training data, then the frequency-based probability estimate will be zero.

This is problematic because it will wipe out all information in the other probabilities when they are multiplied.

A simple numerical correction is to set these zero probabilities to a small value to regularize against this case.



# NAIVE BAYES: APPLICATION AS SPAM FILTER

- In the late 90s, Naive Bayes became popular for e-mail spam filter programs
- Word counts were used as features to detect spam mails (e.g., "Viagra" often occurs in spam mail)
- Independence assumption implies: occurrence of two words in mail is not correlated
- Seems naive ("Viagra" more likely to occur in context with "Buy now" than "flower"), but leads to less required parameters and therefore better generalization, and often works well in practice.

