1 Comparing models

	MLE	**MAP**
Discriminative	• Linear Regression • Logistic Regression • Logistic regression with polynomial features	• Linear regression with L2 regularization • Logistic Regression with Laplace Prior
Generative	• Naive Bayes	• Naive Bayes with Laplace smoothing

Consider the list of models we've learnt so far and place them in the correct boxes above:

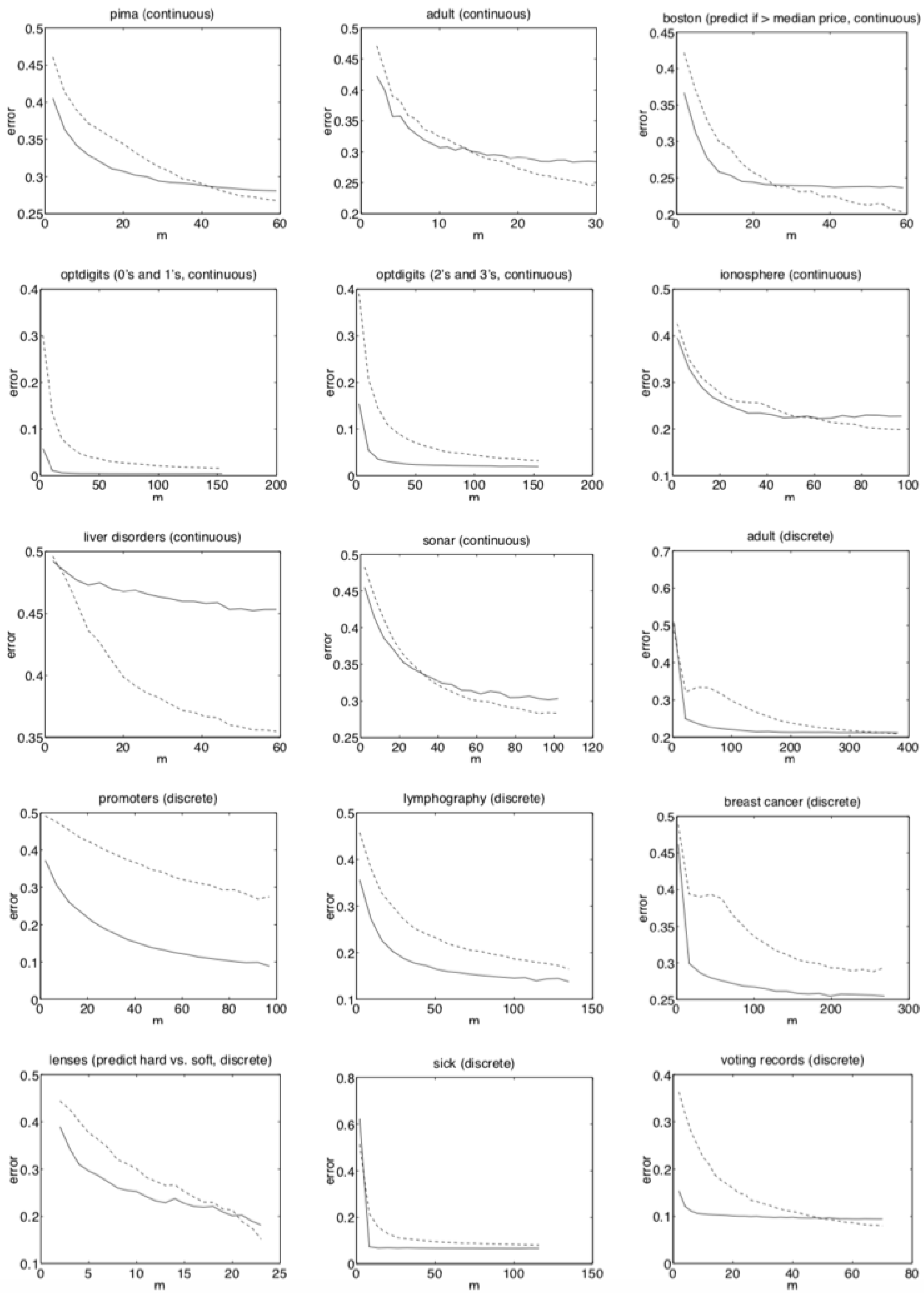
- Linear regression
- Logistic regression
- Linear regression with L2 regularization
- Logistic regression with Laplace prior
- Logistic regression with polynomial features (HW3)
- Naive Bayes
- Naive Bayes with Laplace smoothing

Conceptual Recap: What is Laplace Smoothing?

Naive Bayes takes the product of probabilities. If unobserved words appear in the testing set, probability would then be driven to 0. Laplace smoothing ensures non-zero probabilities by adding pseudo-counts (fake data), such as add-1 or add-λ smoothing.

In context of the SPAM detection exercise done in class, when finding p('some word' | spam=1), we added 1 to the numerator and 2 to the denominator (so assuming a document that contained all words and another document with no words).

2 Performance of Generative vs Discriminative models



These are 15 experiments run by Andrew Ng and Michael Jordan at U.C. Berkeley using 15 standard UCI data sets, in order to compare the performance of Generative vs Discriminative models in real applications. (FYI, this is the link to the original paper: https://ai.stanford.edu/~ang/papers/nips01-discriminativegenerative.pdf) In each of the plots, the X-axis is number of samples, and Y-axis is average error across 1,000 random splits of training/validation set. Look at the graphs and answer the following:

a. The two models considered are Naive Bayes for Generative model and Logistic Regression for Discriminative model. Which model does the dashed line represent? What about the solid line?

   The dashed line represent Logistic Regression. The solid line represents Naive Bayes.

b. According to these results, which model has the better asymptotic performance?

   In most cases, Logistic Regression has the better asymptotic performance.

c. What performance advantages does each type of model possess?

   Generative: better performance when data is scarce. Additional ability to simulate new samples (because it models the joint distribution of $\mathbf{x}$ and y.

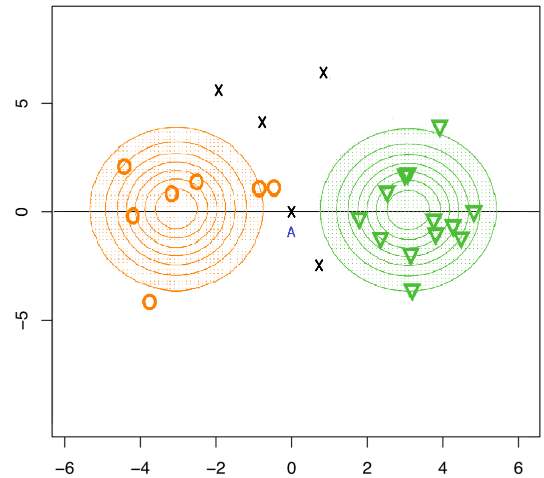
   Discriminative: better asymptotic performance. Better performance on high-dimensional feature space with dependent features.

3 Gaussian Discriminant Analysis (A generative method)

Gaussian discriminant analysis is used when the input features are continuous and $p(\mathbf{x}|y)$ is modeled as a multivariate Gaussian distribution.
Note: Since we're dealing with $p(\mathbf{x}|y)$, it is a generative model, despite its name!

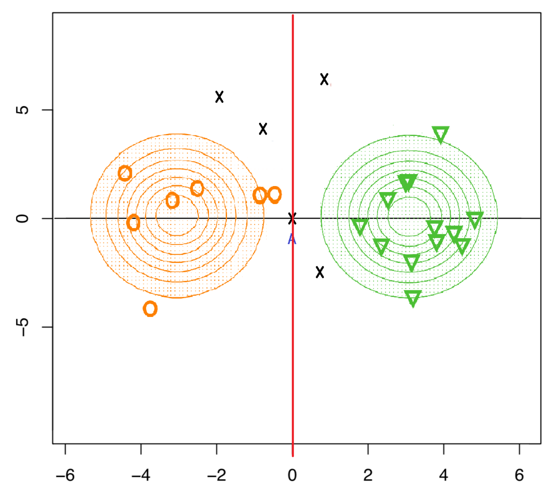
a. Consider two Gaussian distributions as formulated and visualized below:

$\mathbf{x}|y = 0 \sim \mathcal{N}(\mu_{y=0}, \mathbf{I})$
$\mathbf{x}|y = 1 \sim \mathcal{N}(\mu_{y=1}, \mathbf{I})$



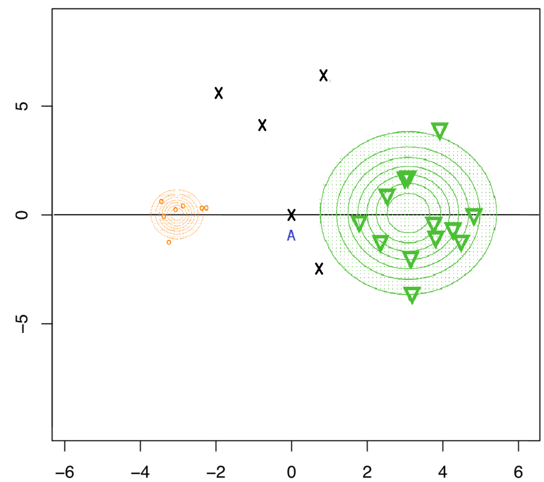
i. X are some observed data points. You are given that point A lies on the midpoint between the two distribution centers. Label each data point with its likely class.
(hint: Both distributions have the same covariance!)

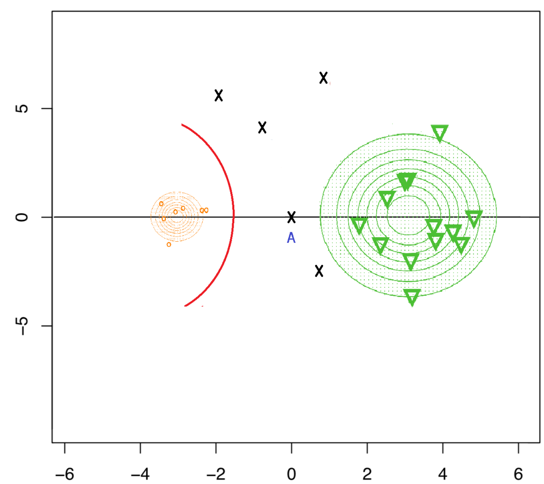
ii. If we were to draw a boundary separating the two distributions, where would the boundary be and what would it look like?



The boundary will be linear.

b. Now supposed a different distrbution, whose covariance matrix is $\frac{1}{5}$I. Re-label the data points again. Point A lies on the midpoint between the two distribution centers. How does the boundary change?





The boundary will now be the zero-crossing of a quadratic function (only a segment of it is drawn above).