154 Midterm Two Concepts

Classification

- response is categorical (discrete)(categories = groups = classes)
- k-class problems (Y $\in$ 1...k)

Logistic Regression (two-class)

- Model: P(Y=1|X=x) = $\frac{e^{x^T\beta}}{1+e^{x^T\beta}}$.
- β is estimated via MLE. (Optimizing: Newton-Raphson) (Iterated reweighted least Squares) (No analytical solution for the β estimate).
- Linear decision boundaries.
- Does not have any normality assumptions.

Multinomial (k-class) Logistic Regression

- Model:

$$log\frac{P(Y=k|X=x)}{P(Y=K|X=x)} = x^T\beta_k$$

Linear Discriminant Analysis (parametric)

- Model:
likelihood:

$$X|Y=k \sim N(\mu_k, \Sigma)$$

prior probabilities:

$$\pi_k = P(Y=k)$$

(Use Bayes rule)
posterior probabilities:

$$P(Y=k|X=x) = \frac{\pi_k f_k(x)}{\sum_{l=1}^{k} \pi_l f_l(x)}$$

f(x): MVN density function (μ_k, Σ_k) at x
- Estimation:
$\hat{\pi_k}$ = obs prop for class k = $\frac{n_k}{n}$
$\hat{\mu_k}$ = obs k^{th} group mean = $\frac{1}{n}\Sigma_{i,y_i=k}X_i$
$\hat{\Sigma}$ = pooled within-group cov matrix = $\frac{n_1-1}{n-K}\hat{\Sigma}_1 + \cdots + \frac{n_K-1}{n-K}\hat{\Sigma}_K$
- linear decision boundaries (discriminant function/ score, check the slides) (To find Decision boundary, $\Delta_k(x) = \Delta_l(x)$ for all k != l)
- require normality assumption
- Assume common within - group covariance matrix

Quadratic Discriminant Analysis (parametric)

- Model:

$$X|Y = k \sim N(\mu_k, \Sigma_k)$$

$$\pi_k = P(Y = k)$$

- quadratic decision boundaries
- require normality assumption
- diff cov matrix for each group

KNN (Non-parametric)

- Idea: Find the k nearest observations in the training data and do a majority vote. (k is a hyper-parameter, tuned via cross validation)
- Model-free! (No distributional assumption on the data)
- Standardization of predictors are highly recommended! (Why? Most distance measures (eg, Euclidean distance) are affected by the scale of predictors. — give large weights to large scale (magnitude) predictors)

Decision Tree (Non-parametric)

Check the slides!

Performance Metric

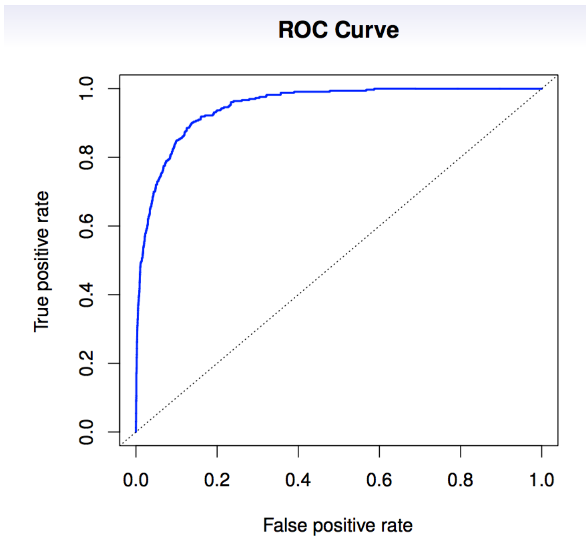
- Accuracy = number of correct predictions / number of predictions. (error/misclassification rate = 1 - accuracy)
- Issues:

1. class imbalance (e.g, 99% of data belong to class 1) (The trivial classifier that predicts all obs to be 1 regardless of the predictors achieves 99% accuracy).
2. Different types of errors might carry a different cost. (Confusion matrix deals with it)

Confusion matrix (K x K): k class, ij-th entry = number of observations s.t. actual class = i and predicted class = j.

Roc Curve

A more comprehensive view of the classifier's performance (without restricting to a particular threshold).



Real Line: generated by changing the thresholds in the classification rules.
Dash Line: theoretical performance of a random classifier.
Perfect classifier is the horizontal line at 1.
TPR = TP / P = sensitivity.
FPR = FP / N = 1 - specificity = 1 - TN / N.
AUC = area under roc curve = prob that a randomly chosen +.instance has higher ranking/score than a randomly chosen -.instance