Introduction to Statistical Learning

November 20, 2020

Classification

- Response variable is qualitative/categorical
- Classification: approach to predict categorical response
- Most widely-used classifiers: *logistic regression*, *linear discriminant analysis*, *K-nearest neighbors*
- Classical examples:
  - Hospital: attribute one of three medical conditions to a new patient
  - Banking: determine if transaction is fraudulent
  - DNA sequence: determine which DNA mutations are deleterious
- Linear regression makes sense for classification problems in binary settings (yes/no: 0 or 1) or when ordering makes sense (mild, moderate, severe). Usually not the case. Otherwise we obtain different regression models based on the encoding of Y.

Logistic Regression

- Logistic regression models the probability that Y belongs to a certain category: $P(Y = i|X)$. We make predictions based on this probality (> 0.5?). Threshold is choosen.
- We assume encoding $Y \in \{0, 1\}$. To model $p(X) = P(Y = 1|X)$, we need a mapping from X to $\{0, 1\}$, to avoid negative values or values > 1. This is the reason why we can't choose a linear mapping of the form $p(X) = \beta_0 + \beta_1 X$.
- *Logistic function* is used: $p(X) = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}$. Gives a S-shaped curve.
- We often consider the *odds* instead of probabilities in this setting: $\frac{p(X)}{1-p(X)} = e^{\beta_0+\beta_1 X}$.

- Estimation of parameters with maximum likelihood instead of least squares. We try to find $\hat{\beta}_0$ and $\hat{\beta}_1$ such that $\hat{p}(x_i)$ is as close as possible to the value of $y_i \in \{0, 1\}$.
- Likelihood function: $\ell(\beta_0, \beta_1) = \prod_{i:y_i=1} p(x_i) \prod_{i':y_{i'}=0}(1 - p(x_{i'}))$.