From sample exam

$t > t_{n-K-1,\alpha/2}$; n is samples, K is predictors, α is confidence

$t = \dfrac{\beta_i - \beta^{null}_{(0)}}{SE(\beta_i)}$

when $\nu \to \infty$, $t_\nu \to Z$, Z is normal

$CI: \beta \pm (t_{n-K-1,\alpha/2}) \cdot SE$; null is n-2 DOF

	True class 1	class 2
predicted c1		
c2		

$RSE = \sqrt{\frac{1}{n-2} RSS}$ (est. of σ of ε)

$TSS = \sum (y_i - \bar{y})^2$

$RSS = \sum (y_i - \hat{y}_i)^2$

Linear Regression

- $Y = \beta_0 + \beta_1 X + \varepsilon$
- $\varepsilon_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$

$SE(\beta_1) = \dfrac{\sigma^2}{\sum_{i=1} (x_i - \bar{x})^2}$

$SE(\beta_0) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2} \right]$

$\sigma^2 = Var(\varepsilon)$

$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{f}(x_i))^2 = \frac{RSS}{n}$

- Least squares!

$\hat{\beta}_1 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \dfrac{\text{sample covariance}}{\text{sample variance}^2}$

$F_{p, n-p-1, \alpha} = \dfrac{(TSS - RSS)/p}{RSS/(n-p-1)}$; p is predictors

$R^2 = \dfrac{TSS - RSS}{TSS}$

- $r = \dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}}$

TSS - RSS = SSR (regression sum of squares)

$SSR = \sum (\hat{y}_i - \bar{y})^2$

Lesson 3

Classification

$p_K(x) = Pr(Y = K | X = x)$

Maximum Likelihood:

$\ell(\beta_0, \beta) = \prod_{i: y_i = 1} p(x_i) \prod_{i: y_i = 0} (1 - p(x_i))$

Choose β_0, β_1 to maximize

$\log_{(ln)} \left(\dfrac{p(X)}{1 - p(X)} \right) = \beta_0 + \beta_1 X$

$P_{Y|X}(y|x) = Pr(Y = y | X = x) = [p(x)]^y [1 - p(x)]^{1-y}$, $y = 0, 1$

Logistic Regression:

$p(X) = \dfrac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}$

Logit: "odds"

$Z = \dfrac{\hat{\beta}_i - \beta^{(null)}}{SE(\hat{\beta}_i)}$ standard normal

Bayes' Theorem: $Pr(Y = K | X = x) = \dfrac{Pr(X = x | Y = K) \cdot Pr(Y = K)}{Pr(X = x)}$

LDA or $Pr(Y = K | X = x) = \dfrac{\pi_K f_K(x)}{\sum_{\ell=1}^{K} \pi_\ell f_\ell(x)}$ ($f_K(x)$ ← density, π_K ← prior)

False positive: Type 1

False negative: Type 2

Sensitivity: TP/P (recall)

Specificity: TN/N

precision: TP/TP+FP

Neg predictive value: TN/TN+FN

F_1 score $= 2 \times \dfrac{precision \times recall}{precision + recall}$

$F_\beta = \dfrac{\beta^2 + 1}{\frac{\beta^2}{recall} + \frac{1}{precision}}$

Naive Bayes: LDA but w/ $f_K(x) = \prod_{j=1}^{p} f_{jK}(x_j)$

QDA

$\delta_K(x) = -\frac{1}{2}(x - \mu_K)^T \Sigma_K^{-1} (x - \mu_K) + \log \pi_K$

LDA w/ p=1 Gaussian Density:

$f_K(x) = \dfrac{1}{\sqrt{2\pi}\sigma_K} e^{-\frac{1}{2}\left(\frac{x - \mu_K}{\sigma_K}\right)^2}$

or maximize $\delta_K(x) = x \cdot \dfrac{\mu_K}{\sigma^2} - \dfrac{\mu_K^2}{2\sigma^2} + \log(\pi_K)$

$\hat{\mu}_K = \dfrac{1}{\pi_K} \sum_{i: y_i = K} x_i$

$\sigma^2 = \left(\dfrac{1}{n - K}\right) \sum_{K=1}^{K} \sum_{i: y_i = K} (x_i - \hat{\mu}_K)^2 = \sum_{K=1}^{K} \dfrac{n_K - 1}{n - K} \cdot \hat{\sigma}_K^2$

$\hat{Pr}(Y = K | X = x) = \dfrac{e^{\delta_K(x)}}{\sum_{\ell=1}^{K} e^{\delta_\ell(x)}}$

Lesson 4

Resampling

$LOOCV = \frac{1}{n} \sum_{i=1}^{n} MSE_i$

$KFCV = \sum_{K=1}^{K} \frac{n_K}{n} MSE_K$

Bootstrapping: $\bar{S}^* = \sum_b S(Z^{*b}) / B$

$\hat{Var}[S(Z)] = \frac{1}{B-1} \sum_{b=1}^{B} (S(Z^{*b}) - \bar{S}^*)^2$

Z^* is a bootstrap dataset

$SE_B(\hat{\alpha}) = \sqrt{\frac{1}{B-1} \sum_{r=1}^{B} (\hat{\alpha}^{*r} - \bar{\hat{\alpha}}^*)^2}$

$\hat{Err}_{boot} = \frac{1}{B} \frac{1}{N} \sum_{b=1}^{B} \sum_{i=1}^{N} L(y_i, \hat{f}^{*b}(x_i))$

Lesson 5: Model Selection + Regularization

- $C_p = \frac{1}{n} (RSS + 2d\hat{\sigma}^2)$; $\hat{\sigma}^2$ is $\hat{var}(\varepsilon)$, d is # parameters
- σ (aka $var(\varepsilon)$) = RSE – hard to know

- $AIC = -2 \log L + 2 \cdot d$ (may like) or $AIC = \frac{1}{n\hat{\sigma}^2} (RSS + 2d\hat{\sigma}^2)$
- $BIC = \frac{1}{n} (RSS + \log(n) d\hat{\sigma}^2)$

- Adjusted $R^2 = 1 - \dfrac{RSS/(n-d-1)}{TSS/(n-1)}$

Ridge Regression:

$RSS = \sum_{i=1} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2$

RR estimates should minimize: $RSS + \lambda \sum_{j=1}^{p} \beta_j^2$

First standardize via: $\tilde{x}_{ij} = \dfrac{x_{ij} - \bar{x}_j}{\sqrt{\frac{1}{n} \sum_{i=1} (x_{ij} - \bar{x}_j)^2}}$

The Lasso: $RSS + \lambda \sum_{j=1}^{p} |\beta_j|$

Elastic net for redundant variables, $L_1 + L_2$.

$RSS + \lambda \left[\frac{1}{2}(1 - \alpha) \|\beta\|_2^2 + \alpha \|\beta\|_1 \right]$

Dimension Reduction

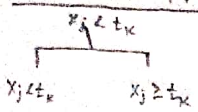
$Z_m = \sum_{j=1}^{p} \phi_{mj} X_j$; $y_i = \theta_0 + \sum_{m=1}^{M} \theta_m z_{im} + \varepsilon$

$\beta_j = \sum_{m=1}^{M} \theta_m \phi_{jm}$; $\theta \to \beta$, $Z \to X$, Z via ϕ

Choosing ϕ w/ PCR/PCA.

PLS considers Y too

Lesson 6: Tree Based Methods



- Find boxes that minimize

$$RSS = \sum_{j=1}^{J} \sum_{i \in R_j} (y_i - \hat{y}_{R_j})^2 \quad (+ \alpha |T|)$$ pruning

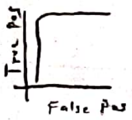
Gini index: $G = \sum_{K=1}^{K} \hat{p}_{mk}(1 - \hat{p}_{mk})$ (classification) "purity"

A measure of total variance across k classes.

or

Cross-entropy: $D = -\sum_{K=1}^{K} \hat{p}_{mk} \log(\hat{p}_{mk})$

Bayes Error Rate: $1 - E(\max_j Pr(Y=j \mid X))$



- SMOTE: up + down sampling to account for imbalanced data.
- $MSE = Var(\hat{f}(x_0)) + Bias(f(x_0)) + Var(\varepsilon)$
- KNN: $p_j(x_0) = \frac{1}{K} \sum_{i \in N_0} I(y_i = j)$