FreeFellow

FORMULA SHEET

Exam SRM

SOA · Statistics for Risk Modeling

21
FORMULAS

5
TOPICS

freefellow.org/soa-srm/formulas

BASICS OF STATISTICAL LEARNING

2 items

Bias-variance tradeoff

$E[(y-\hat{f}(x))^2] = \text{Bias}^2(\hat{f}) + \text{Var}(\hat{f}) + \sigma_\varepsilon^2$

Irreducible error σ_ε^2 cannot be reduced

k-fold cross-validation error

$CV_{(k)} = \frac{1}{k}\sum_{j=1}^{k} \text{MSE}_j$

Each fold serves as validation once

LINEAR MODELS

9 items

F-statistic (regression)

$F = \frac{(SS_{\text{tot}} - SS_{\text{res}})/p}{SS_{\text{res}}/(n-p-1)}$

Tests H_0 : all slope coefficients are zero

LASSO penalty

Minimize: $\sum(y_i - \hat{y}_i)^2 + \lambda\sum_{j=1}^{p}|\beta_j|$

L_1 penalty; produces sparse solutions (exact zeros)

Ridge regression penalty

Minimize: $\sum(y_i - \hat{y}_i)^2 + \lambda\sum_{j=1}^{p}\beta_j^2$

L_2 penalty; shrinks but does not zero out coefficients

Elastic net penalty

Minimize: $\sum(y_i - \hat{y}_i)^2 + \lambda_1\sum|\beta_j| + \lambda_2\sum\beta_j^2$

Combines LASSO (L_1) and Ridge (L_2)

R-squared

$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}} = 1 - \frac{\sum(y_i-\hat{y}_i)^2}{\sum(y_i-\bar{y})^2}$

Variance inflation factor (VIF)

$\text{VIF}_j = \frac{1}{1-R_j^2}$

R_j^2=R^2 from regressing X_j on all other predictors

VIF>5–10 indicates multicollinearity

Adjusted R-squared

$\bar{R}^2 = 1 - \frac{(1-R^2)(n-1)}{n-p-1}$

p=number of predictors (excludes intercept)

BIC

$BIC = p\ln n - 2\ln\hat{L}$

Penalizes complexity more than AIC for $n > 7$

AIC

$AIC = 2p - 2\ln\hat{L}$

p=number of parameters; lower is better

TIME SERIES MODELS

5 items

AR(1) model

$X_t = \phi X_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim WN(0,\sigma^2)$

Stationary iff $|\phi| < 1$

Ljung-Box test statistic

$Q = n(n+2)\sum_{k=1}^{m}\frac{\hat{\rho}_k^2}{n-k}$

Tests H_0 : first m autocorrelations are zero

Distributed $\chi^2(m)$ under H_0

ACF of AR(1)

$\rho(h) = \phi^h, \quad h = 0,1,2,\ldots$

Decays geometrically; PACF cuts off after lag 1

ARMA(1,1) model

$X_t = \phi X_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$

Stationary iff $|\phi| < 1$

MA(1) model

$X_t = \varepsilon_t + \theta\varepsilon_{t-1}, \quad \varepsilon_t \sim WN(0,\sigma^2)$

Always stationary

DECISION TREES

3 items

Gini impurity

$G = \sum_{k=1}^{K} \hat{p}_k(1 - \hat{p}_k) = 1 - \sum_{k=1}^{K} \hat{p}_k^2$

$\hat{p}_k$=fraction of class k in node

Bagging (bootstrap aggregation)

Train B trees on bootstrap samples; aggregate predictions

$\hat{f}(x) = \frac{1}{B} \sum_{b=1}^{B} \hat{f}_b(x)$

Reduces variance without increasing bias

Entropy (node impurity)

$H = -\sum_{k=1}^{K} \hat{p}_k \log \hat{p}_k$

UNSUPERVISED LEARNING TECHNIQUES

2 items

K-means objective function

$\min_{C_1,\ldots,C_K} \sum_{k=1}^{K} \sum_{i \in C_k} \|x_i - \bar{x}_k\|^2$

$\bar{x}_k$=centroid of cluster k

PCA — proportion of variance explained

$\mathrm{PVE}_k = \frac{\lambda_k}{\sum_{j=1}^{p} \lambda_j}$

λ_k=kth eigenvalue of covariance (or correlation) matrix