

Examinable Topics for Revision

ACTL3142 & ACTL5110 Statistical Machine Learning for Risk Applications



Statistical Learning

- What is statistical (machine) learning?
- prediction and inference
- interpretability and flexibility
- supervised and unsupervised learning
- regression and classification problems
- mean squared error
- bias-variance trade-off
- K-nearest neighbors



Linear Regression

- Linear Regression
- Simple Linear Regression
- Assumptions of the Model
- Least Squares Estimates (LSE)
- LSE Properties & Uncertainty
- Maximum Likelihood Estimates
- Confidence Intervals
- Multiple Linear Regression
- Qualitative predictors
- Linear Algebra / Matrix Approach
- Residual standard error (RSE)
- R-squared (R^2)
- Linear model as an approximation
- Random error ϵ
- dummy variables
- Interaction terms ($X_i X_j$)
- Quadratic terms (X_i^2)



Linear Regression II

- Linear Model Selection
 - Subset selection
 - Indirect methods
- Subset Selection
 - Best subset
 - Forward stepwise
 - Backwards stepwise
 - Hybrid stepwise
- C_p , AIC, BIC, and Adjusted R^2
- Problems with Linear Regression
 - Non-linearity of response-predictor relationships
 - Correlation of error terms
 - Non-constant variance of error terms
 - Outliers, high-leverage points, collinearity, and confounding effects



Logistic Regression

- binary classification
- Logistic regression
- odds, log-odds
- logits
- error rate (misclassification)
- confusion matrix
- thresholds
- ROC curve
- AUC
- Poisson regression



Generalised Linear Models

- generalised linear models
- link functions
- distributions
- exponential family
- mean function
- variance function
- canonical link functions
- null model
- full model
- deviance & scaled deviance
- residuals
- degrees of freedom



Cross-validation and regularisation

- Train, validation & test sets
 - Validation Set Approach
 - k -fold Cross-Validation
 - Leave-One-Out Cross-Validation
 - Test error
- Regularisation (shrinkage)
 - ridge
 - lasso
 - ~~elastic net~~



Moving Beyond Linearity

- interpolation & extrapolation
- polynomial regression
 - monomials
 - orthogonal polynomials
- step functions
 - basis function expansion
 - piecewise polynomial functions
- regression splines
 - knots
 - natural splines
 - cubic splines
- smoothing splines
- local regression
- generalised additive models (GAMs)



Tree-based Methods

- decision trees
- classification trees
- regression trees
- pruning
- ensemble methods
- bagging
- random forests
- boosting
- ~~gradient boosting~~



Unsupervised Learning

- k -means clustering
- hierarchical clustering
- principal components analysis

