STAT406- Methods of Statistical Learning Lecture 25

Matias Salibian-Barrera

UBC - Sep / Dec 2017

Review

- Response variable Y, vector of explanatory variables $\mathbf{X} = (X_1, X_2, \ldots, X_p)'$
- Interest in predicting Y or understanding the relationship
- Optimal predictor for Y based on $\mathbf{X}$ using a squared loss function?

Review

- The regression function

$$f(\mathbf{X}) = E[Y|\mathbf{X}]$$

- Model-based estimates (e.g. linear or polynomial regression)
- Feature selection can be important – prediction power negatively affected by highly correlated explanatory variables; sometimes the interest is in identifying **all** important features.

Review

- Estimate *future* prediction error:

$$E_{(Y,\mathbf{x})}\left[\left(Y-\hat{f}_n(\mathbf{X})\right)^2\middle|\mathcal{D}_n\right]$$

- "Prediction error on the training set" (aka "residual sum of squares")?
- Cross-validation?
- AIC?, Mallow's C_p?

Review

- Regularization
- Ridge regression, LASSO, Elastic Net
- All have tuning parameters that need to be chosen
- CV? AIC?
- How many “parameters” does a regularized estimator have?
- EDF

Review

- What if the regression function

$$f(\mathbf{X}) = E[Y|\mathbf{X}]$$

  is not linear in $\mathbf{X}$?
- Model-based: non-linear regression
- Polynomial, splines approximations; kernel (local) regression estimators
- Regression trees
- Why trees if we have kernel regression or splines?

Review

- Classification - another prediction problem
- Optimal classifier using a 0-1 loss function?
- Need to estimate

$$P(Y = g|\ \mathbf{X})$$

for each class label g and each $\mathbf{X}$

Review

- Model-based estimators
- Given a model for the distribution of $\mathbf{X}|Y = g$ we can obtain

$$P(Y = g|\ \mathbf{X}) \propto f(\mathbf{X}|\ Y = g)\, P(Y = g)$$

- Model-free estimators
- Nearest neighbours
- Logistic / multinomial
- Trees

Review

- Ensembles work better than individual predictors / classifiers
- Bagging is one way to build an ensemble, but variability of the combined prediction can be a concern
- Random Forests
- Out-of-bag prediction error estimates, feature importances
- Boosting

Review

- Unsupervised learning, “no response”
- K-means, Dissimilarities (divisive)
- Model-based clustering, EM-algorithm (divisive)
- Hierarchical clustering (agglomerative)
- Principal Components, dimension reduction, “best” lower dimensional approximations