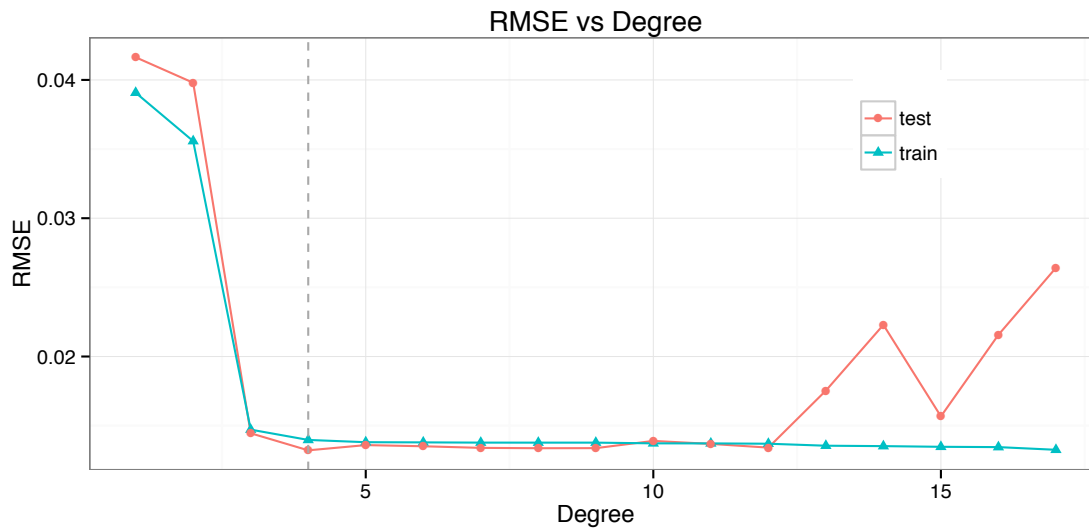


APMA-E4990: Homework 2

Problem 1. Cross-validation for polynomial regression

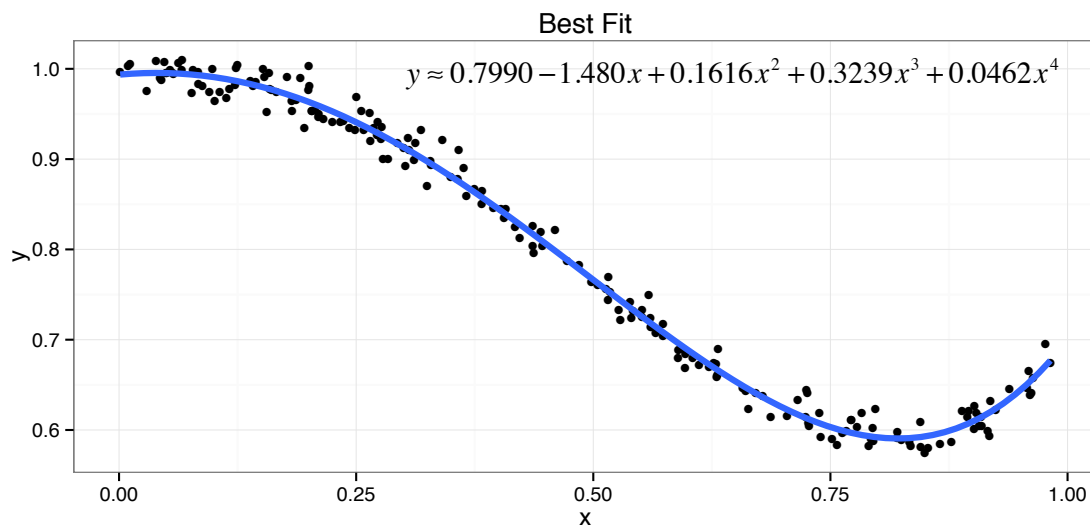
Root-mean-square error (RMSE) vs Polynomial degree

A 4th degree polynomial minimizes the RMSE on the testing data.



Best fit polynomial

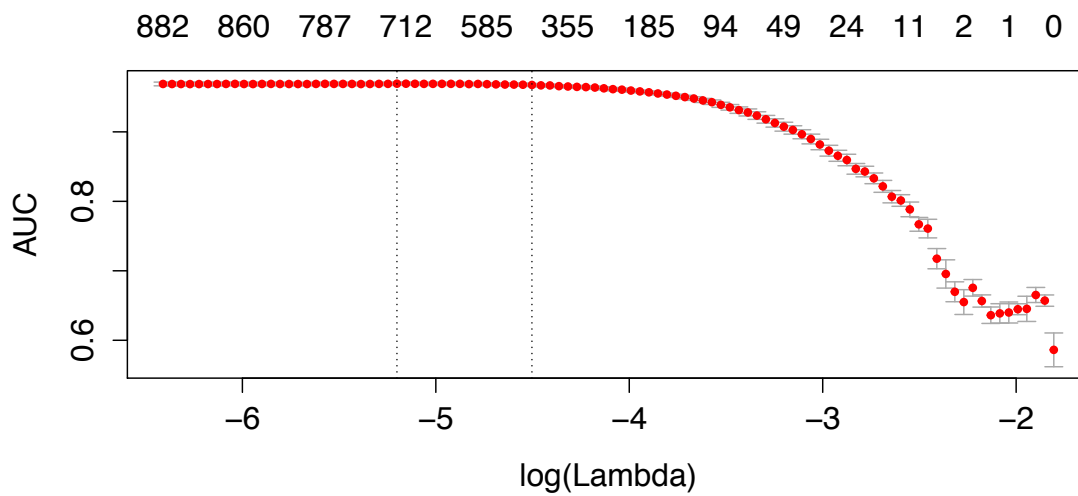
Full data set with the coefficients and plot of the 4th degree polynomial overlaid.



Problem 2. Logistic regression for article classification

Cross-validation curve

AUC vs $\log(\text{Lambda})$



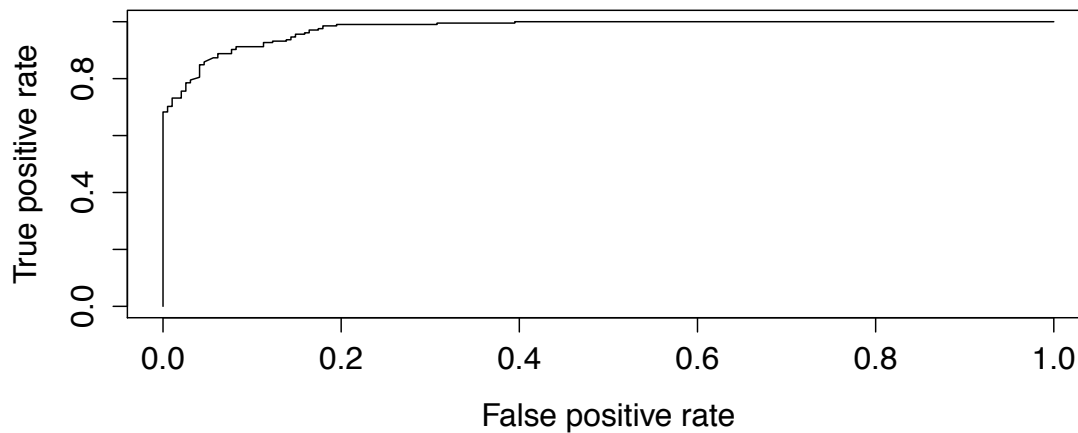
Accuracy and AUC

Accuracy: 0.905

AUC: 0.9767

ROC curve for the testing data

A 4th degree polynomial, indicated with a dashed gray line, minimizes the RMSE on the testing data.



Informative words

Words indicative of a Business article

weight	word
2.266917	obamas
1.877835	collect
1.738303	nasty
1.673156	updated
1.396554	nbc
1.210733	company
1.178025	kevin
1.126458	publishing
1.094364	(Intercept)
1.081952	executive
1.073323	also

Words indicative of a World article

weight	word
-3.112892	burial
-2.672924	died
-2.655990	thailand
-2.563773	holy
-2.410290	anticorruption
-2.395889	canadian
-2.383248	abuja
-2.360692	videosharing
-2.291570	islands
-2.245923	minister