Homework 1

Daniel Tshiani

2025-05-20

1

a

a flexible model would preform better because the sample size is large so the chances of overfitting decreases. also the number of predictors is small so the model can focus on the few variables non-linear relationships

b

an inflexible model would be better because of the small number of observations. overfitting can happpend rather quickly. and with the large number of predictors, it would be better to generalize them which an inflexible model does.

c

a flexible model would be better because it would be able to better caputure local max/mins and the changes in slope.

d

a inflexible model would perform better because a flexible model would be fitting the noise. so the training data might be well fit but the testing data wouldn't be a good fit.

2

a

this is a regression problem. This a inference problem. n would be the 500 firms. P or the independent variables would be profit, number of employees, and industry salary. the dependent variable would be the CEO salary.

b

this is a classification and prediction problem. N is 20 similar products that were previously launched. P is success or failure of previously launched products, price charged for the product, marketing budget, competition price, and ten other variables.

c

this is a classification and prediction problem. n is the weekly data for all 2012. the predictors are the precentage change in the dollar, the percentage change in the british market, and the german market.

3

a

- a prediction classification model can be estimating players risk of injury. so they would be at a high risk of injury or a low risk of injury. predictors could include minuted played in the last 2 weeks, number of sprints, age.

-predicting match outcome would be a classification problem. predictors could include home/away status, formation used, recent team form.

b

- if we want to estimate the market value of a soccer player a regression model would be useful. that would be a prediction problem. the dependent variable would be their market value and the independent variables could include how many goals they scored, games played, assists, passing percentage.

- if we wanted to see if market value is effected by red cards a regression model would also work. this would be a inference problem. the dependent variable would be market value again and the dependent variable would be number of red cards.

c

- grouping players by style of play could be a cluster analysis problem. predictors could include postion, dribbles completed, passes completed, shots, tackles. this would be a prediction problem.

- grouping players youth players for training programs could be done via cluseter analysis. this would be a prediciton problem. predictors could include age, strength test like bench press, skill level, sprint speed.