Darius nyaundi 763-01 Machine Learning

4. What are the regression coefficients (thetas)?

coef

Intercept  88.6581

Gender[T.'Male']  -1.4794

Location[T.'St. Mary's Medical Center'] -0.8565

Location[T.'VA Hospital']  -1.7348

SelfAssessedHealthStatus[T.'Fair'] -2.7510

SelfAssessedHealthStatus[T.'Good'] 0.5864

SelfAssessedHealthStatus[T.'Poor'] 0.4593

Age 0.0803

Height 0.4696

Weight -0.0134

Smoker 9.6731

5. How do you interpret those numbers?

-It shows how the independent variable influences the dependent variables.

-The definition says “coefficients represent the mean change in the response variable for one unit of change in the predictor variable while holding other predictors in the model constant.” The similarity will be the slopes.

-You can build a relationship. For example:

Intercept + coef Age\*#Age = Systolic blood pressure

88.6581 + 0.0803\* #Age= Systolic blood pressure

Intercept + coef Height\*#Height = Systolic blood pressure

88.6581 + 0.4696\* #Height= Systolic blood pressure

Intercept + coef Weight\*#Weight = Systolic blood pressure

88.6581 + -0.0134\* #Weight= Systolic blood pressure

6. If you need to identify one outlier record, which record is a potential outlier? How do you reach this conclusion?

Record 92. Because when we detect the leverage we can find that record 92 has a extreme predictor for X values.

7. If you need to identify one or few useless features (independent variables or predictors), which one(s) will you choose? Why do you reach this conclusion?

A useless feature I believe, is the Location independent variable. The location has nothing really to do with an individual’s systolic blood pressure. It can show the results of individual’s systolic blood pressure at that location, but it doesn’t determine the factors as to why the individuals systolic blood pressure is the level that its at.