DSFA
Spring 2018

Lecture 38

Multiple Regression

Announcements

- lab10 release delayed

Prediction

Techniques so far:

- Linear regression: predict quantitative variable using one quantitative variable
- Nearest neighbor: predict categorical variable using several quantitative variables

Today: predict quantitative variable using several quantitative variables

(Demo)

Multiple Linear Regression

With multiple attributes *x1, x2, …, xn*, predict

$$y = a1*x1 + a2*x2 + … + an*xn + b$$

where weights *a1, a2, …, an* and *b* minimize RMSE

If there's just one attribute, simplifies to

$$y = a*x + b$$

(Demo)

Minimize RMSE

With multiple attributes *x1, x2, …, xn*, predict

$$y = a1*x1 + a2*x2 + … + an*xn + b$$

where weights *a1, a2, …, an* and *b* **minimize RMSE**

Minimization

minimize(f)

    returns the x value that minimizes f(x)

minimize(f, array=True, start=a)

    returns the array x that minimizes f(x)

    where a is an array of the same length as x

(Demo)

Nearest Neighbor Regression

- Given new point *(x1, x2, …, xn)*,
- Find k nearest neighbors to that point, each with a *y* value
- Predict the mean of those *y* values

(Demo)

Also Nearest Neighbor Regression