Fitting linear models by least Squares (ISLR Ch. 3)

Given: training data $(x_1, y_1) \ldots (x_n, y_n)$

Goal: • Make prediction rule $f(x)$ that predicts value of Y for query x.

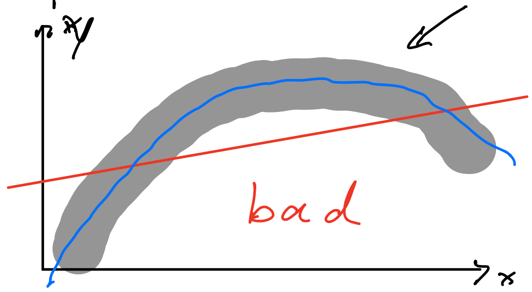
• Summarize relationship between X and Y.

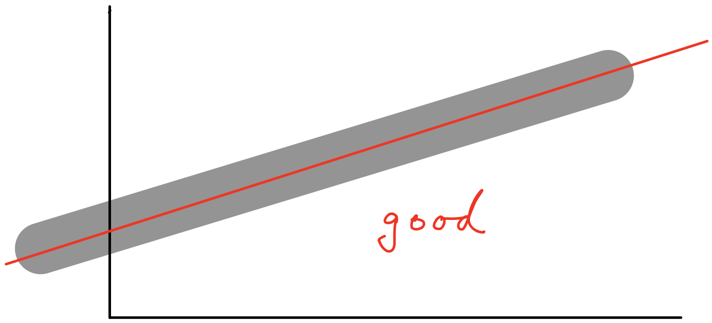
We know: If we want to minimize expected squared prediction error, the optimal rule is $f(x) = E(Y|x)$. We estimate f from the training sample.

We will now talk about fitting a straight line $\ell(x) = b_0 + b_1 x$. Why

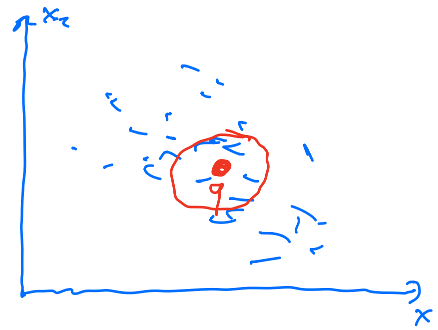
- If it looks like $E(Y|x)$ is linear then straight line gives parsimonious summary of association.

- Straight line may have better predictive performance.





Straight line "technology" can be generalized to problems with multiple predictors and non-linear associations



Two predictors

$$\ell(x_1, x_2) = b_0 + b_1 x_1 + b_2 x_2$$

Nonlinear associations

$$\ell(x) = b_0 + b_1 x + b_2 x^2$$

Back to straigh line fitting

$$\ell(x) = b_0 + b_1 x$$

How to choose b_0, b_1

Choose b_0, b_1 to minimize resubstitution error

$$(\hat{b}_0, \hat{b}_1) = \underset{b_0, b_1}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2 \quad (*)$$

choose b_0, b_1 to give best predictive performance for training sample

By differentiating wrt b_0, b_1 we get

$$\hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x}$$

$$\hat{b}_1 = \sum (x_i - \bar{x})(y_i - \bar{y}) / \sum (x_i - \bar{x})^2$$