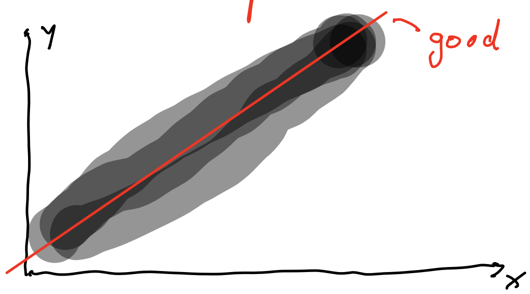
Expansion based prediction methods

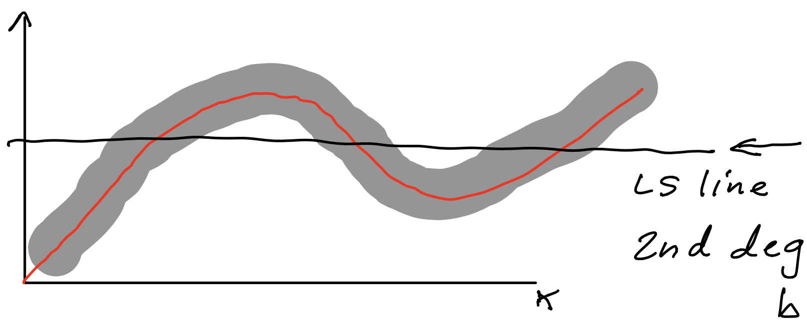
Contrast with local averaging

Local averaging is a "nonparametric" prediction method - it makes only weak assumptions about the nature of dependence between x and y.

In contrast to fitting a straigh line by least squares which assumes that $E(Y|x) = b_0 + b_1 x$

Consider simplest case $p=1$





LS line bad!

2nd degree polynomial bad!

3rd degree probably ok

What to do?

- Pick a collection or dictionary of of "basis functions" $B_1(x) \dots B_K(x)$ for which (hopefully)

$$E(Y|X) \approx \sum_{i=1}^{K} a_i \cdot B_i(x)$$

- Find $\hat{\underline{a}} = \underset{\underline{a}}{\operatorname{argmin}} \|\underline{y} - X \underline{a}\|^2$

  where $X_{ij} = B_j(x_i)$

  The i-th row of X is $(B_1(x_i) \dots B_K(x_i))$

  The j-th column of X is $(B_j(x_1) \dots B_j(x_n))^T$

  BTW: Always assume that one of the basis functions is the constant

Examples for dictionaries

- Polynomials: $B_i(x) = x^i \quad i = 1 \dots K-1$

  $B_K(x) = 1$

  $\sum a_i B_i(x) =$ polynomial of degree $K-1$

  order K

- Piecewise constant functions

  Pick "Knot" positions $c_1 \leq c_2 \leq c_{K-1}$

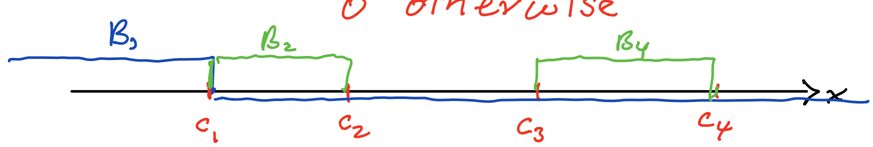
  Define $c_0 = -\infty \qquad c_K = +\infty$

  Define $B_i = I((c_{i-1}, c_i], x) \quad i = 1 \dots K$

  $\shortparallel$

  1 if $x \in [c_{i-1}, c_i)$

  0 otherwise



How to place the Knots ?

Placing Knots at all x_i no good
would interpolate

Better suggestion
Pick 10% percentile, 20% percentile, ets