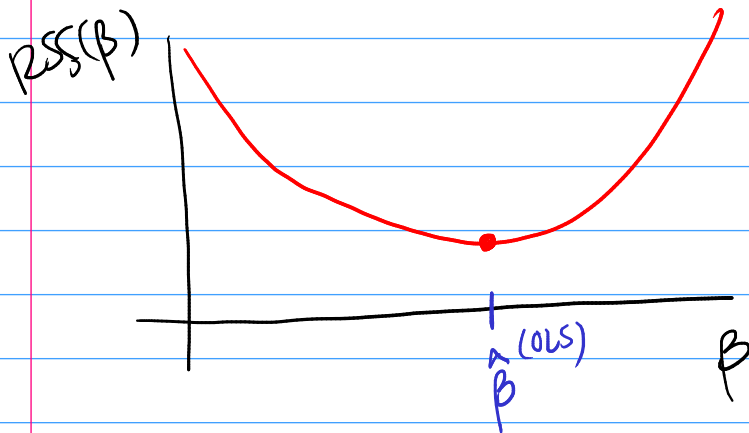
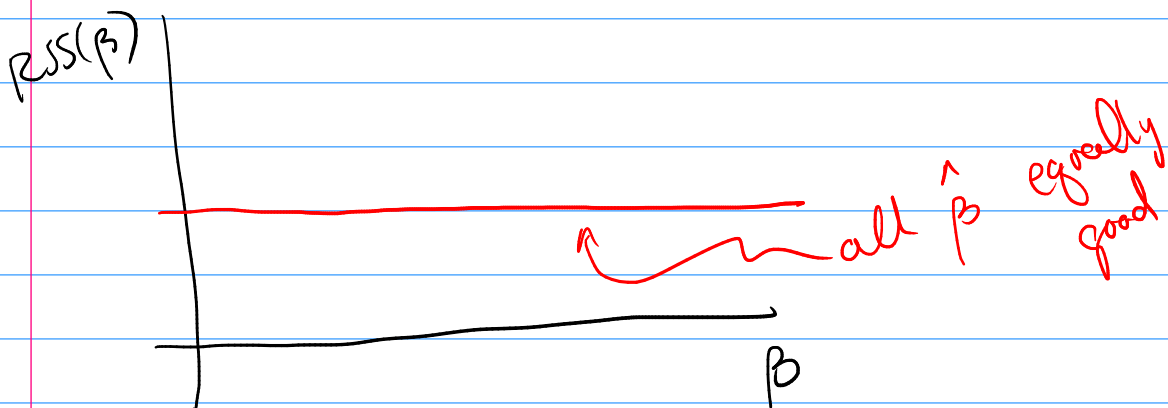


Lecture 4:

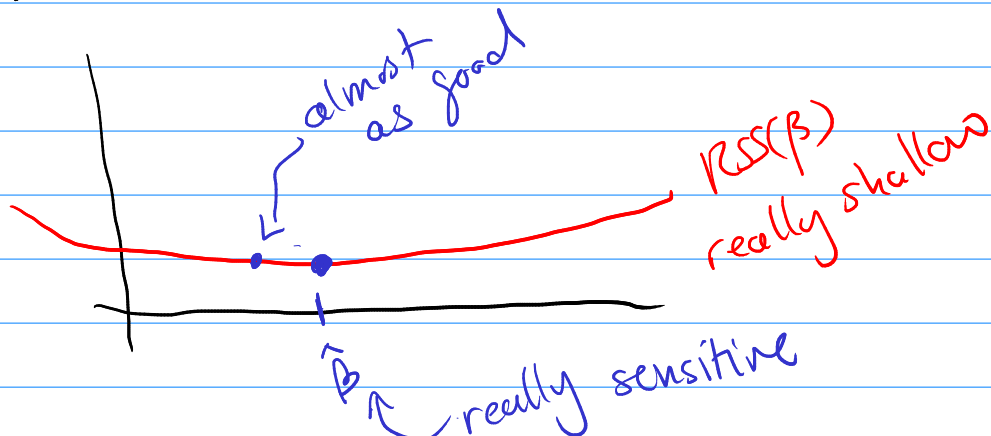
① $X^T X$ is invertible



② $X^T X$ not invertible



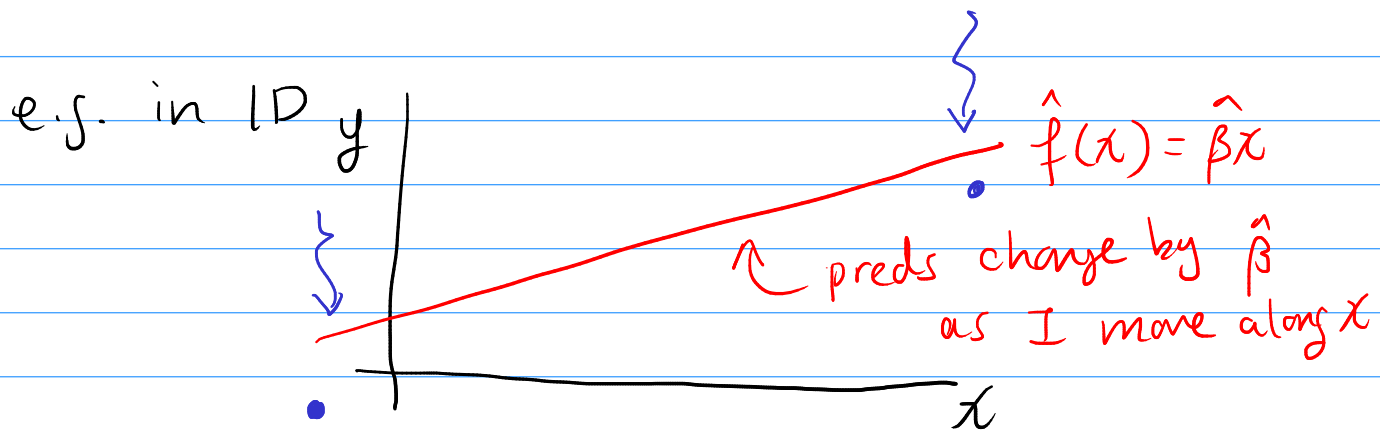
③ $X^T X$ close to not invertible



kNN Regression (k - nearest neighbors regression)

Recall for LR: we have a really strong global assumption about f

$$f(\underline{x}) = \underline{x}^T \beta$$



Furthermore: training data affects the fit far away

Benefit: strong global assumption make \hat{f} practical to find

kNN makes a weaker local assumption that values of $\hat{f}(\underline{x})$ only depend on nearby training points.

$\rightarrow k$ = integer that determines how many "nearby" training points we consider

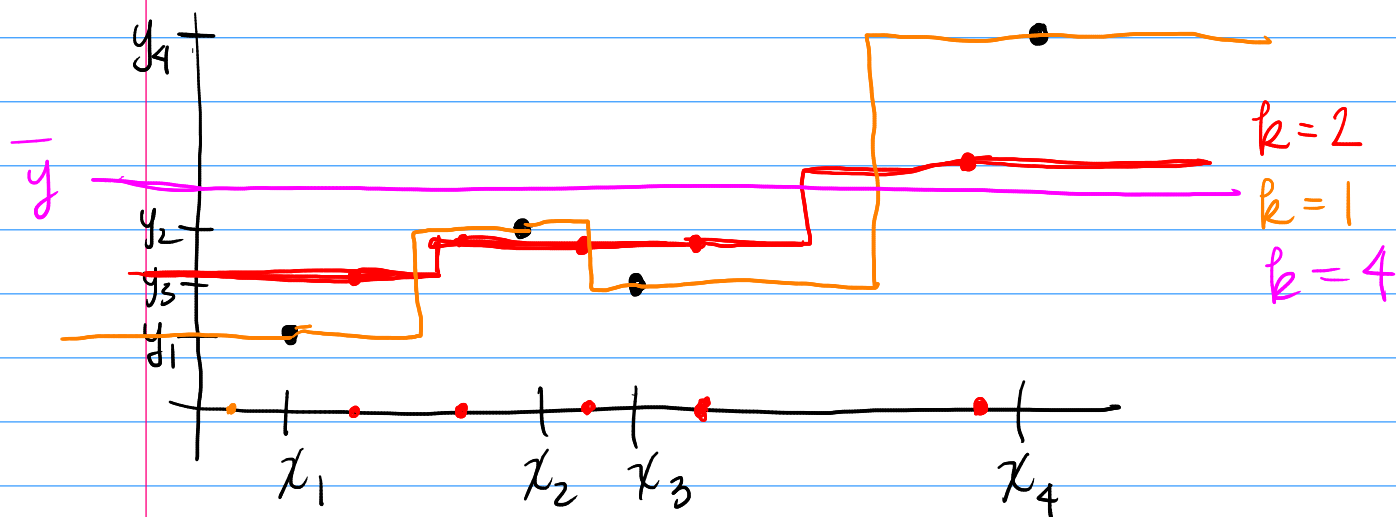
Given training data $\{(\underline{x}_n, y_n)\}_{n=1}^N$

kNN fits \hat{f} as

$$\hat{f}(\underline{x}) = \frac{1}{k} \sum_{n \in N_k(\underline{x})} y_n = \text{arg. } y_n \text{ over } k \text{ nearest neighbors to } \underline{x}$$

$N_k(\underline{x})$ = k neighborhood of \underline{x}
= indices of k nearest training pts to \underline{x}

for numeric typically use euclidean



What happens as we change k ?

general rule: k controls the flexibility of the method
↳ how complicated of a function \hat{f} is

Small $k \rightarrow$ very flexible method
eg. $k=1$ interpolates data

large $k \rightarrow$ very inflexible method
eg. $k=N$ I just predict \bar{y}

Comparison w/ OLS:

OLS reduces \mathcal{F} to a p -dim'l space.

k NN reduces \mathcal{F} to a N/k dim'l space.

as $k \uparrow$ I reduce
the dim of the space

(VC dimension
Vapnik-Chervonensk)

converse: $k \downarrow$ I increase the dim