Lecture 25

Can generalize to regression easily w/ a squared error loss.

More difficult to generalize to other losses

Can do <u>Gradient Boosting</u>.

Shrinkage:

$0 < v < 1$

$$\hat{f}(x) = \text{Sign}\left(\sum_{b=1}^{B} v \alpha_b \hat{f}_b(x) \right)$$

Kernel Regression

For regression wanted to model $E[Y|X=x]$

Ways to do this:

global

(1) linear regression

$$E[Y|X=x] \approx x^T\beta$$

(2) polynomial regression

$$E[Y|X=x] \approx \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots \beta_p x^p$$

local

(3) <u>KNN</u>

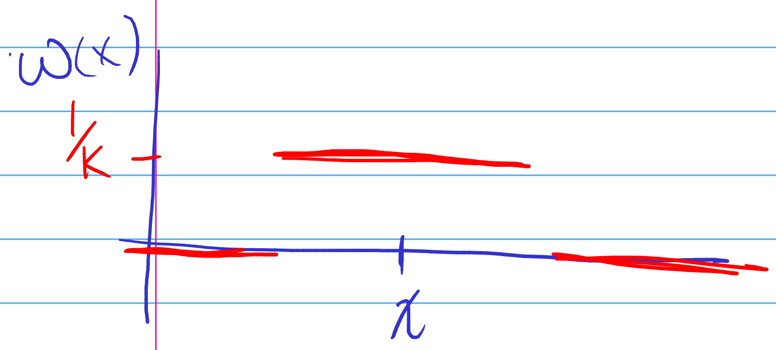
$$E[Y|X=x] = \text{mean of nearby } y\text{s}$$

KNN regression:

$$\hat{f}(x) = \sum_{n=1}^{N} w_n(x) y_n$$

$$w_n(x) = \frac{1}{K} \mathbb{1}(x_n \in N_K(x))$$

$$= \begin{cases} 1/K & x_n \in N_K(x) \\ 0 & \text{else} \end{cases}$$



Note: weight isn't very smooth - maybe use smoother weights?

This is called Kernel Smoothing (regression)

Defn: a kernel is a function K

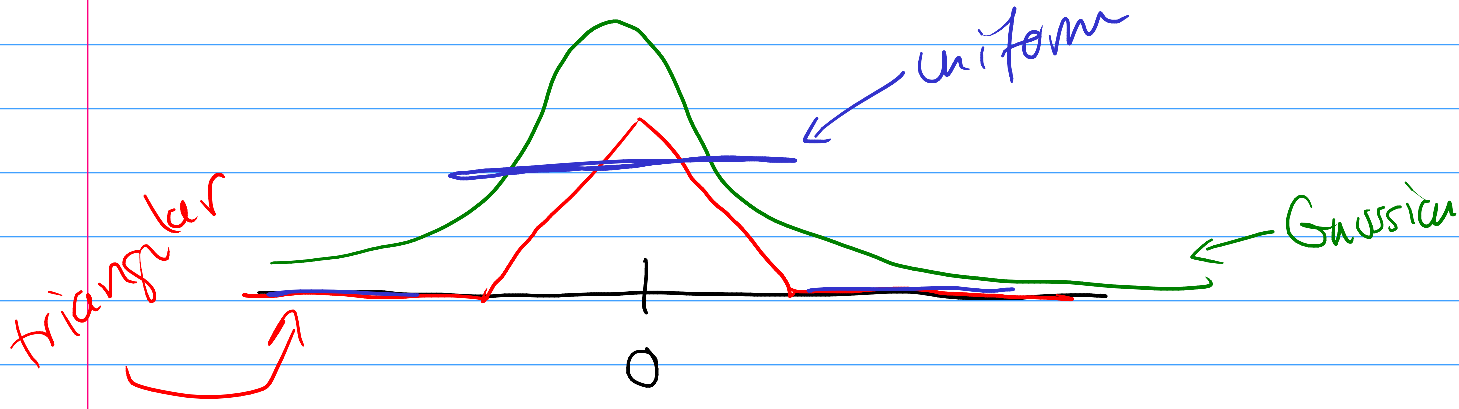
1. $K : \mathbb{R}^p \to \mathbb{R}$
2. $K(x) \geq 0$

typically also assume

3. $\int_{\mathbb{R}^p} K(x)\,dx = 1$ [PDF of some RV]
4. $K(x) = K(-x)$ (symmetric)

Ex. $K(x) = e^{-x^2}$

$K(x) = \frac{1}{1+x^2}$



Kernels can be used to define "closeness"

Ex. $K(r) = e^{-r^2}$

then $K(x-y) = e^{-(x-y)^2}$

Kernel Smoother:

$$\hat{f}(x) = \sum_{n=1}^{N} w_n(x) y_n$$

where $w_n(x) \propto \underbrace{K(x - x_n)}_{\text{closeness of } x \text{ to } x_n}$

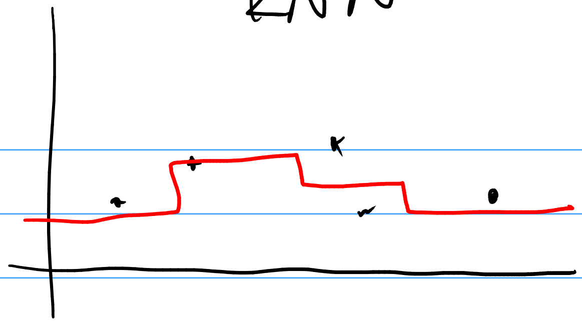
typ. want $\sum_n w_n(x) = 1$

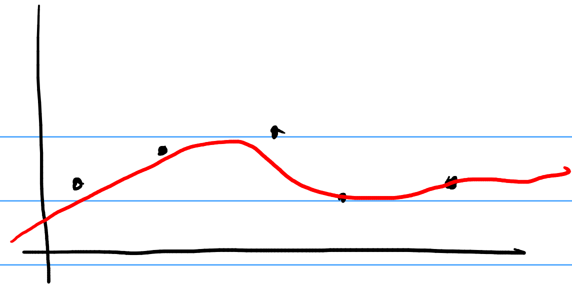
so we can defn

$$w_n(x) = K(x - x_n) \Big/ \sum_m K(x - x_m)$$

Known as Nadaraya-Watson Est. (NW)

KNN





Kernels typ. have some "bandwidth" parameter that controls "closeness"

Ex

$$K_\gamma(x) = \exp(-\gamma x^2)$$

$$K_\gamma(x) = \frac{1}{1 + x^2/\gamma}$$

tuning gamma is the hard work for fitting.