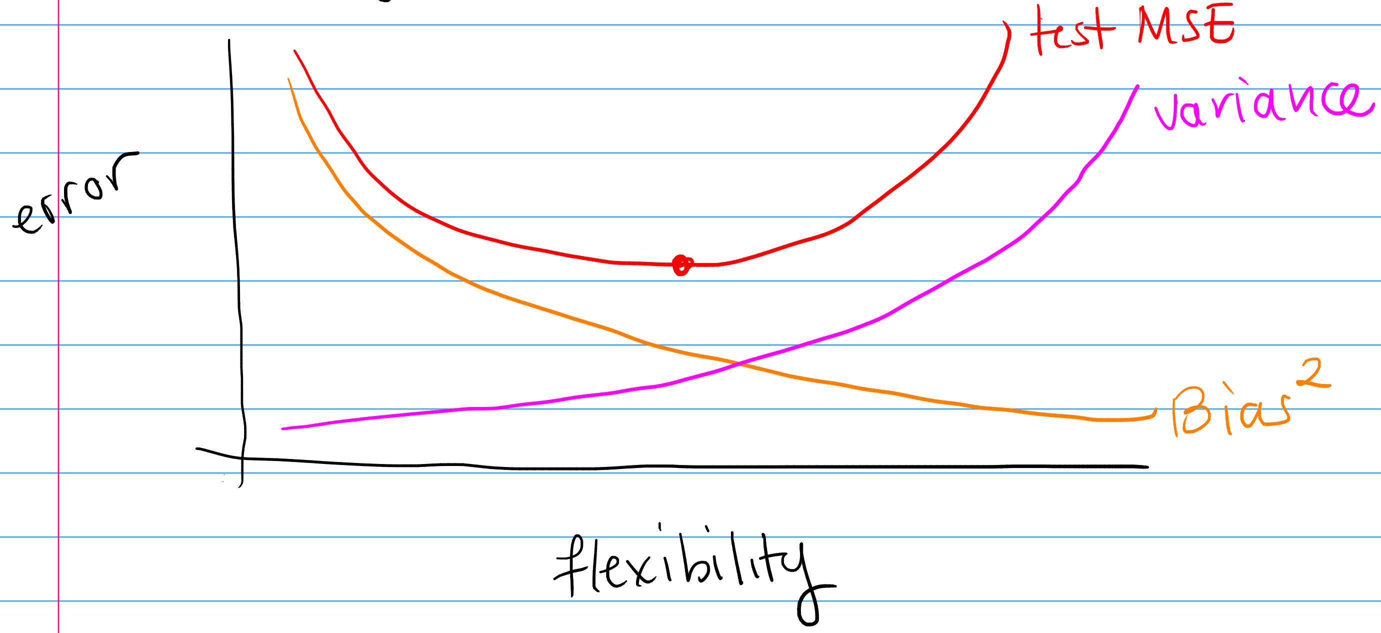
Lecture 8: Risk Minimization

Mathematically: $MSE = Bias^2 + Var$ + Irreducible



Mathematically:

$$Bias(\hat{f}(\mathbf{x}_0)) = E[\hat{f}(\mathbf{x}_0)] - f(\mathbf{x}_0)$$

$$Var(\hat{f}(\mathbf{x}_0)) = E[(\hat{f}(\mathbf{x}_0) - E[\hat{f}(\mathbf{x}_0)])^2]$$

MSE

$$Err(\mathbf{x}_0) = E[(Y_0 - \hat{f}(\mathbf{X}_0))^2]$$

$$= E[(Y_0 - E[\hat{f}(\mathbf{x}_0)])^2] \quad ①$$

$$+ E[(\hat{f}(\mathbf{x}_0) - E[\hat{f}(\mathbf{x}_0)])^2] \quad ②$$

$$+ 2E[(Y_0 - E\hat{f}(\mathbf{x}_0))(\hat{f}(\mathbf{x}_0) - E[\hat{f}(\mathbf{x}_0)])] \quad ③$$

$$Y_0 = f(\mathbf{x}_0) + \varepsilon_0$$

① $= \mathbb{E}[(\underbrace{f(x_0)}_{a} + \overbrace{\varepsilon_0}^{b} \underbrace{- \mathbb{E}[\hat{f}(x_0)]}_{a})^2] = \mathbb{E}[(a+b)^2]$

$= \cancel{\mathbb{E}}[(\underbrace{f(x_0) - \mathbb{E}[\hat{f}(x_0)]}_{\text{deterministic}})^2] = \text{Bias}^2$

$+ \mathbb{E}[\varepsilon_0^2] = \sigma^2$

$+ 2\mathbb{E}[\varepsilon_0(\underbrace{f(x_0) - \mathbb{E}[\hat{f}(x_0)]}_{\text{deterministic}})] = 0$

$\mathbb{E}[\varepsilon_0] = 0$

$\Rightarrow \text{Bias}^2 + \sigma^2$

② $= \mathrm{Var}(\hat{f}(x_0))$, $\mathrm{Var}(z) = \mathbb{E}[(z - \mathbb{E}z)^2]$

$z = \hat{f}(x_0)$

③ $\mathbb{E}[(Y_0 - \mathbb{E}[\hat{f}(x_0)])(\hat{f}(x_0) - \mathbb{E}[\hat{f}(x_0)])]$

independent b/c $Y_0 \perp\!\!\!\perp$ training data

$A \perp\!\!\!\perp B \Rightarrow \mathbb{E}[AB] = \mathbb{E}[A]\mathbb{E}[B]$

$\downarrow$

$\mathbb{E}[Y_0 - \mathbb{E}[\hat{f}(x_0)]]\,\mathbb{E}[\underbrace{\hat{f}(x_0) - \mathbb{E}[\hat{f}(x_0)]}_{0}] = 0$

$z = \hat{f}(x_0)$

$\mathbb{E}[z - \mathbb{E}[z]]$

$= \mathbb{E}[z] - \cancel{\mathbb{E}}\,\mathbb{E}[z] = 0$

Q: what's the best $\hat{f}$ theoretically?

L = loss (prev. called M)

$L(Y, f(x)) = (Y - f(x))^2$ = sq. loss

$L(Y, f(x)) = |Y - f(x)|$ = abs. loss

$$f^* = \underset{f}{\operatorname{argmin}}\ \underbrace{\mathbb{E}[L(Y, f(\underset{\sim}{x}))]}_{\text{Risk}}$$

Turns out can actually get an answer.

Let $(\underset{\sim}{X}, Y) \sim P$ ← joint dist

then

$\mathbb{E}[L(Y, f(\underset{\sim}{x}))]$

$$= \underbrace{\mathbb{E}_{\underset{\sim}{X}}\, \underbrace{\mathbb{E}[L(Y, f(\underset{\sim}{x})) \mid \underset{\sim}{X} = \underset{\sim}{x}]}_{A(\underset{\sim}{x})}}$$

← want to minimize

Iterated Expectation

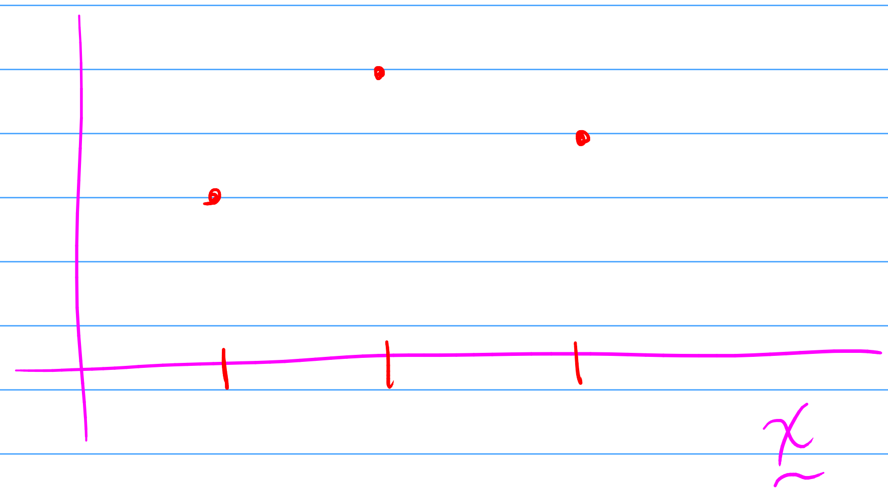
$\mathbb{E}[A] = \mathbb{E}_B \mathbb{E}[A|B]$

$$= \int A(\underset{\sim}{x})\, p(\underset{\sim}{x})\, d\underset{\sim}{x}$$

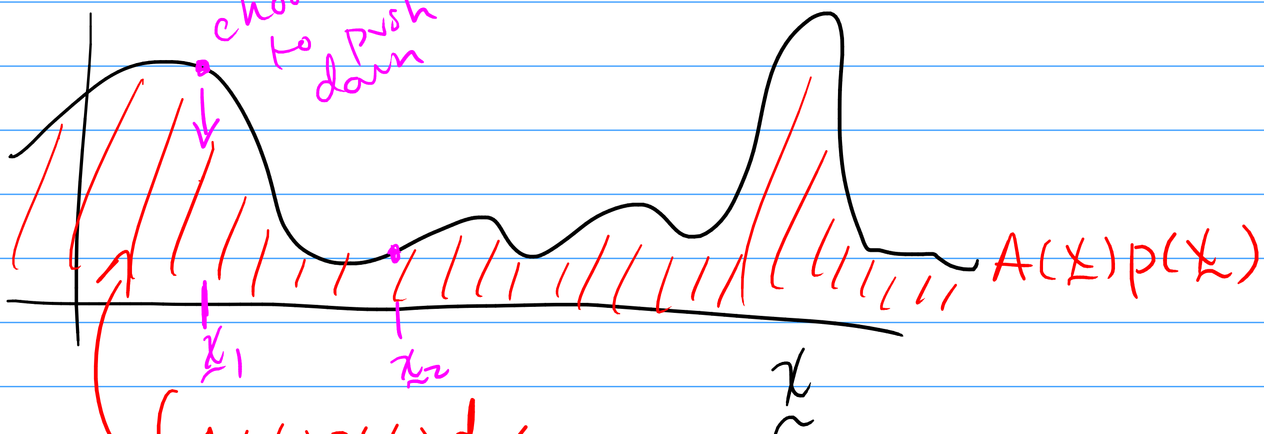
← need to choose f to minimize

depends on f

I need to choose f, can do it seperately for each $\underset{\sim}{x}$.



Look at $A(\underset{\sim}{x})p(\underset{\sim}{x})$



$$\text{area} = \int A(\underset{\sim}{x})p(\underset{\sim}{x})\,d\underset{\sim}{x}$$

because $p(\underset{\sim}{x})$ doesn't depend on $f(\underset{\sim}{x})$

only need to try to minimize $A(\underset{\sim}{x})$

So, really just need to choose $f(\underset{\sim}{x})$ at each $\underset{\sim}{x}$ to minimize $A(\underset{\sim}{x})$

i.e.

$$f^*(\underset{\sim}{x}) = \underset{f(\underset{\sim}{x})}{\operatorname{argmin}}\ \mathbb{E}\left[L(Y, f(\underset{\sim}{x})) \mid \underset{\sim}{X} = \underset{\sim}{x}\right]$$

$$= \underset{c}{\operatorname{argmin}}\ \mathbb{E}\left[L(Y, c) \mid \underset{\sim}{X} = \underset{\sim}{x}\right]$$

more concrete, L = sq loss

$$f^*(\underset{\sim}{x}) = \underset{c}{\operatorname{argmin}}\ \mathbb{E}\left[(Y-c)^2 \mid \underset{\sim}{X} = \underset{\sim}{x}\right]$$

Aside:

$$\underset{c}{\operatorname{argmin}}\ \mathbb{E}\left[(z-c)^2\right] = \mathbb{E}[z]$$

pf $\mathbb{E}\left[(z-c)^2\right] = \mathbb{E}\left[z^2 + c^2 - 2zc\right]$

$$= \mathbb{E}[z^2] + c^2 - 2c\,\mathbb{E}[z]$$

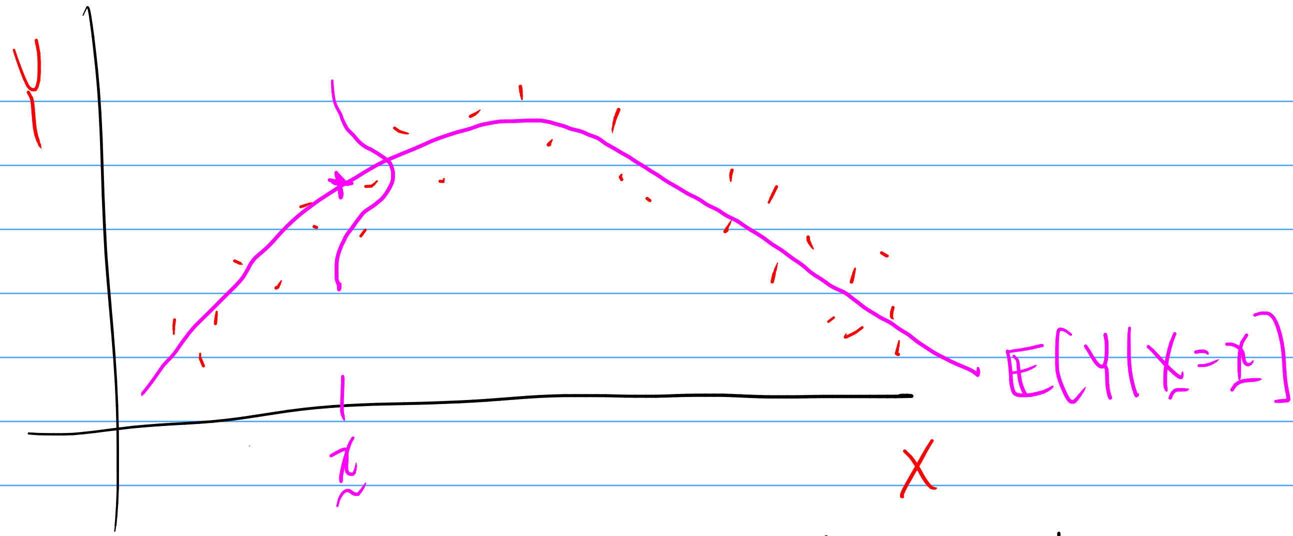
$$\frac{\partial}{\partial c}(\cdots) = 2c - 2\,\mathbb{E}[z] = 0$$

Solve for c

$$\Rightarrow c = \mathbb{E}[z].$$

So

$$f^*(\underset{\sim}{x}) = \mathbb{E}\left[Y \mid \underset{\sim}{X} = \underset{\sim}{x}\right]$$



What about $L(Y, f(x)) = |Y - f(x)|$?

$$f^*(x) = \text{Median}(Y \mid X = x)$$

Problem: I don't know underlying dists.

Can try to approx
$$\hat{f}(\underset{\sim}{x}) \approx \mathbb{E}[Y \mid \underset{\sim}{X} = \underset{\sim}{x}]$$
using my training data.

Ex. $\hat{f}(\underset{\sim}{x}) =$ avg. of y_n's for $\underset{\sim}{x}_n$'s near $\underset{\sim}{x}$

KNN regression.

Ex Make some assumptions about structure of $\mathbb{E}[Y \mid \underset{\sim}{X} = \underset{\sim}{x}]$

in particular assume $\mathbb{E}[Y \mid \underset{\sim}{X} = \underset{\sim}{x}] = \underset{\sim}{x}^T \beta$

OLS
linear regression.

Outline:

- Stat. learning
  - supervised
    - regression
      done some
    - classification
      Now
  - unsupervised

Setup: Classification

$\underline{x} \in \mathbb{R}^p, \quad y \in \mathcal{C}$

$\mathcal{C} = \{c_1, c_2, \dots, c_K\}$

↑ set of possible classes

Goal: find some $\hat{f}$ so that $y \approx \hat{f}(\underline{x})$.

Loss function for classification: 0-1 loss

$$L(Y, \hat{f}(\underline{x})) = \mathbb{1}(Y \neq \hat{f}(\underline{x}))$$

$$= \begin{cases} 0 & \hat{f}(\underline{x}) = Y \\ 1 & \hat{f}(\underline{x}) \neq Y \end{cases}$$

What is f^*?

$$f^*(\underset{\sim}{x}) = \underset{c}{\arg\min}\ \mathbb{E}[L(Y,c) \mid \underset{\sim}{X} = \underset{\sim}{x}]$$

$$= \underset{c}{\arg\min}\ \mathbb{E}[\mathbb{1}(Y \neq c) \mid \underset{\sim}{X} = \underset{\sim}{x}]$$

Claim: $\mathbb{E}[\mathbb{1}(\text{statement})] = \mathbb{P}(\text{statement})$

$$= \underset{c}{\arg\min}\ \mathbb{P}(Y \neq c \mid \underset{\sim}{X} = \underset{\sim}{x})$$

$$= \underset{c}{\arg\min}\ 1 - \mathbb{P}(Y = c \mid \underset{\sim}{X} = \underset{\sim}{x})$$

$$\boxed{f^*(\underset{\sim}{x}) = \underset{c}{\arg\max}\ \mathbb{P}(Y = c \mid \underset{\sim}{X} = \underset{\sim}{x})}$$

Bayes' classifier

Ex: 3-class problem: Cats, Dogs, People

$P(Y = \text{cat} \mid \underset{\sim}{X})$, $P(Y = \text{dog} \mid \underset{\sim}{X})$, $P(Y = \text{people} \mid \underset{\sim}{X})$

predict class w/ largest prob.