SIS - III

Conceptual questions:

1.

- Binary classification: water on/off, spam/not spam
- Multiclass: blood type (A/B/0/...); digits (0,1,2,3,...)
- Regression: predict weather / price for something

2.

y_i	1	1	0	0
$\hat{y}_i$	0.9	0.4	0.3	0.6

(using base-10 logs)

Data: $(y, \hat{y}) = (1, 0.9), (1, 0.4), (0, 0.3), (0, 0.6)$

Formula: $L(\theta) = -\frac{1}{n}\sum_{i=1}^{n}[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)]$

$= \frac{1}{4}[-(1 \cdot \ln(0.9) + 0 \cdot \ln(0.1)) + (-(\ln(0.4) + 0 \cdot \ln(0.6))) +$

$+(-(0 \cdot \ln(0.3) + 1 \cdot \ln(0.7))) + (-(0 \cdot \ln(0.6) + 1 \cdot \ln(0.4)))] =$

$= \frac{1}{4}(-\ln(0.9) - \ln(0.4) - \ln(0.7) - \ln(0.4)) = \frac{2.2946}{4} \approx 0.5737$ nats

3 we notice that the model is <u>overfitting</u>

4. b 5. b 6. c

7.

Error rate $= 1 - \frac{n_i}{n}$, its minimal when l is the majority class: $l = \arg\max_k \cdot n_k$.

Minimum error: $\gamma = \frac{\max_k n_k}{n}$

8. $f(x_i) = c$

$$J(c) = \frac{1}{n}\sum_{i=1}^{n}(y_i - c)^2$$

$$\frac{dJ}{dc} = \frac{1}{n}\sum_{i=1}^{n} 2(c - y_i) = \frac{2}{n}\left(nc - \sum_{i=1}^{n} y_i\right)$$

1st.

$$\frac{2}{n}\left(nc - \sum_{i=1}^{n} y_i\right) = 0$$

$$nc = \sum_{i=1}^{n} y_i$$

$$c = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{y}$$

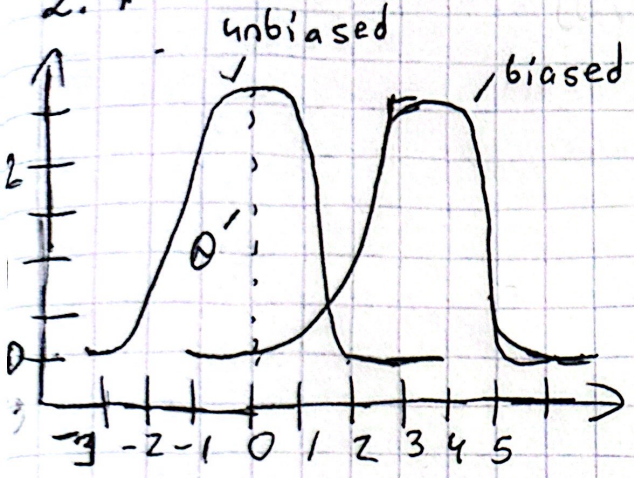
2nd

$$\frac{d^2J}{dc^2} = \frac{2}{n}\sum_{i=1}^{n} 1 = \frac{2}{n} \cdot n = 2 > 0$$

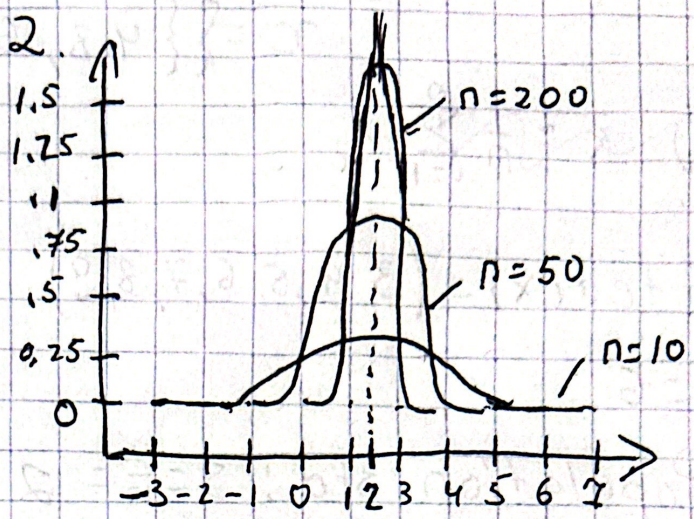
So, $c = \bar{y} = \frac{1}{n}\sum_{i=0}^{n} y_i$ //

Properties of an Estimator

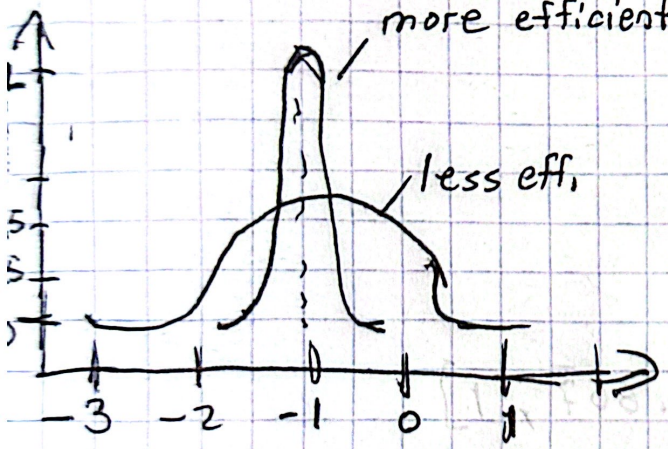
2.1



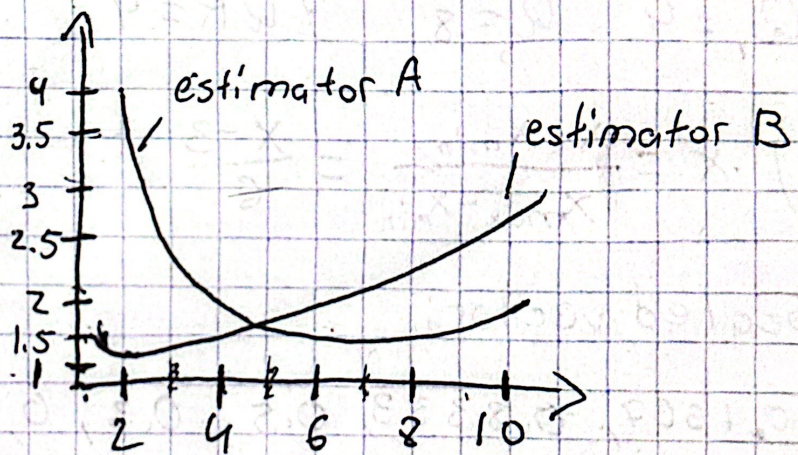
2.2.



2.3



2.5



2.4.

~~$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$~~

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$$

$$= E[(\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta)^2]$$

$$= \underbrace{E[(\hat{\theta} - E(\hat{\theta}))^2]}_{Var(\hat{\theta})} + \underbrace{(E(\hat{\theta}) - \theta)^2}_{Bias(\hat{\theta},\theta)^2} + 2(E(\hat{\theta}) - \theta)\underbrace{E(\hat{\theta} - E(\hat{\theta}))}_{0}$$

$$= Var(\hat{\theta}) + Bias(\hat{\theta},\theta)^2$$

$Var(\hat{\theta})$ — systematic error

how much estimator fluctuates across diff. samples

Data Scaling;

$$x = \{4, 8, 6, 5, 3, 7, 9\}$$

~~$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$~~

Sorted $(x) = (3, 4, 5, 6, 7, 8, 9)$

$\bar{x} = 6$

Population std: $S = \frac{1}{n} = 2$

Median $= 6$

$Q_1 = 4 \quad Q_3 = 8 \quad IQR = 4$

a) $x' = \frac{x - x_{min}}{x_{max} - x_{min}} = \frac{x-3}{6}$

Scaled vector;

$[0,1667, 0.8333, 0.5, 0.3, 0, 0.667, 1]$

mean $= \frac{\bar{x} - 3}{6} = 0,5$

std $= \frac{2}{6} \approx 0,\overline{3}$

$IQR = \frac{4}{6} \approx 0,6667$

b) $x' = \frac{x - \mu}{\sigma} = \frac{x-6}{2}$

Scaled Vector: $[-1, 1, 0, -0.5, -1.5, 0.5, 1.5]$

mean $= 0 \quad$ std $= 1 \quad IQR = 2$

c) $x' = \frac{x - \text{median}}{IQR} = \frac{x-6}{4}$

Scaled Vector:

$[-0,5,\ 0.5,\ 0,\ -0.25,\ -0.75,\ 0.25,\ 0.75]$

median = 0, mean = 0

std = $\frac{2}{4} = 0,5$ IQR = 1

4) Min-Max: preserves order and maps s.min → 0, max → 1

Standard: min → 0, std → 1

Robust: median → 0, IQR → 1, robust to outliers

5.a) Scaling before splitting leaks test information

b) The model can indirectly "see" the test dist.

c) there is optimistic bios in metrics because leakage make data easier.