Exercise 1

part 1

a) Overfitting occurs when a model learns the training data too well, capturing noise & fluctuations rather than the underlying pattern, leading to poor performance on new data. It is detected when the training error is very low, but the validation/test error is high

b) * Bias is the error introduced by approximating a real-world problem with a simplified model. High bias causes underfitting.

* Variance refers to the model's sensitivity to flutuations in the training set. High variance causes overfitting.

c) Regularization adds a penalty term to the loss function to discourage complex Models. Its purpose is to reduce overfitting & variance, improving generalization.

d) Precision is more important. In spam detection, a FP is much worse than FN. We want to be very sure as what we classify as spam is actual spam.

part 2

a) True

b) True

Exercise 2:

1.) $H(Y) = -\sum p(x) \log_2(p(x)) = -\left(\frac{3}{4}\log(\frac{3}{4}) + \frac{1}{4}\log(\frac{1}{4})\right) \approx \underline{0.811}$

* Splitting on X_1:

- $X_1 = 0$ (rows 1,2): $Y = \{0,1\}$. $H(Y|X_1=0) = -\left(\frac{1}{2}\log_2(\frac{1}{2}) + \frac{1}{2}\log_2(\frac{1}{2})\right) = \underline{1}$
- $X_1 = 1$ (rows 3,4): $Y = \{1,1\}$. $H(Y|X_1=1) = \underline{0}$ (pure Node)

$\Rightarrow E(X_1) = \frac{2}{4}(1) + \frac{2}{4}(0) = 0.5$

$\Rightarrow IG(X_1) = 0.811 - 0.5 = 0.311$

* Splitting on X_2:

// the data is symmetric, calculations will be the same

$\Rightarrow IG(X_2) = \underline{0.311}$

2) we can choose randomly. Bothe reduce entropy by the exact same amount.

Excercise 3:

Binary cross-entropy loss formula: $-\frac{1}{N}\sum[y_i \ln(\hat{y}_i) + (1-y_i)\ln(1-\hat{y}_i)]$

1) . sample 1: $-\ln(0.9) \simeq 0.105$

. sample 2: $-\ln(1-0.2) \simeq 0.223$

. sample 3: $-\ln(0.7) \simeq 0.357$

. sample 4: $-\ln(1-0.6) \simeq 0.916$

$\Rightarrow \frac{\Sigma \dots}{N} = \frac{1.604}{4} \simeq \underline{0.4}$

2) 1: $0.9 \geq 0.5 \rightarrow 1 \Rightarrow$ correct

2: $0.2 \geq 0.5 \rightarrow 0 \Rightarrow$ correct

3: $0.7 \geq 0.5 \rightarrow 1 \Rightarrow$ correct

4: $0.6 \geq 0.5 \rightarrow 1 \Rightarrow$ incorrect !

3)

	predicted 0	predicted 1
Actual 0	1 (TN)	1 (FP)
Actual 1	0 (FN)	2 (TP)

Accuracy $= \frac{TP+TN}{Total} = 0.75$

Precision $= \frac{TP}{TP+FP} \simeq 0.67$

Recall $= \frac{TP}{TP+FN} = 1.0$

F1-Score $= 2 \times \frac{Prec \times Rec}{Prec + Rec} = 0.8$

4) Lowering the threshold make it easier to predict "1" (positive).

$\Rightarrow$ Recall stays 1.0 (or increase), Precision will likely decreases.

Exercise 4:

a) * Expected prediction : Mean of predictions = 0.95

* Bias^2: $(\text{Expected predictions} - \text{True})^2 = (1-1)^2 = 0$

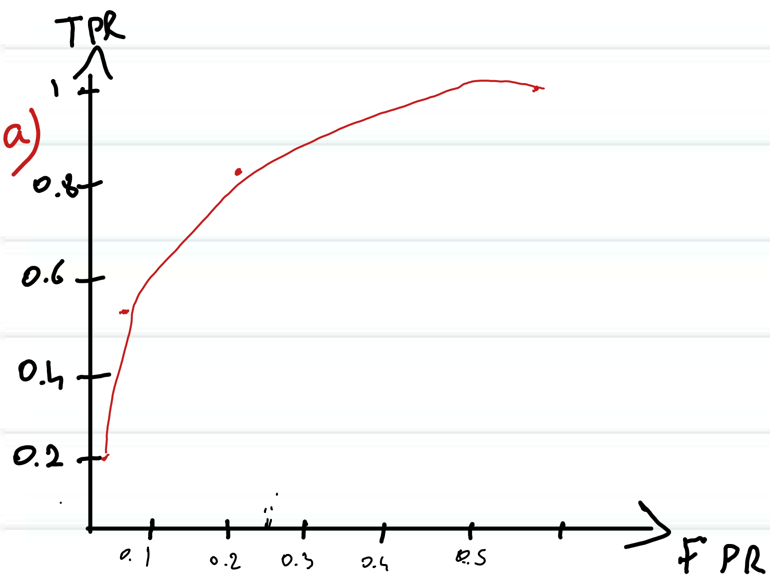
* Variance $\simeq 0.000344$

Total Error = Bias^2 + Variance + Irreducible

$\simeq 0.1038$

Exercise 5:

a)



b) FPR is smaller for threshold 0.5 while keeping TPR $\geq$ 0.8.

c) We cannot tolerate False Negatives. we need to maximize recall (TPR), we choose threshold 0.3.