DS and AI

Machine Learning

Machine Learning 1500+ series

Lecture 02

By - Siddharth Sabharwal Sir

Recap of Previous Lecture

Topic

Topic

Topic

Topic

Machine Learning

Topics to be Covered

o1 Machine Learning

o2

o3

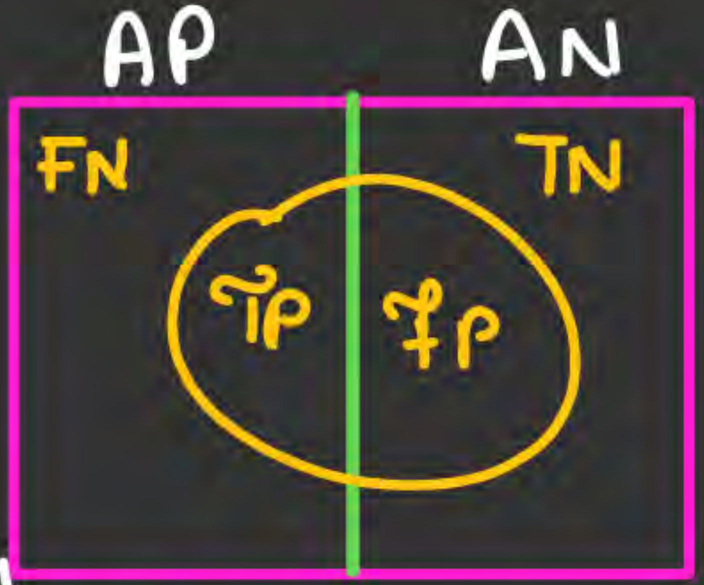
o4

- Logistic Regression :-

$$\Rightarrow \left(\log_e(\text{odd}) = \beta_0 + \beta_1 x\right)$$

$$\Rightarrow P = \left(\frac{1}{1+e^{-(\beta_0+\beta_1 x)}}\right)$$

Confusion matrix



Actual

Pred	P	N
P	TP	FP
N	FN	TN

$$\textbf{Precision} = \boxed{\frac{\text{True Positive}}{\text{Actual Results}}} \quad \text{or} \quad \boxed{\frac{\text{True Positive}}{\text{True Positive + False Positive}}}$$

$$\textbf{Recall} = \boxed{\frac{\text{True Positive}}{\text{Predicted Results}}} \quad \text{or} \quad \boxed{\frac{\text{True Positive}}{\text{True Positive + False Negative}}}$$

$$\textbf{Accuracy} = \boxed{\frac{\text{True Positive + True Negative}}{\text{Total}}}$$

TPR : Recall

$$\text{sensitivity} = \frac{\text{number of true positives}}{\text{number of true positives} + \text{number of false negatives}}$$

$$\boxed{\text{specificity} = \frac{\text{number of true negatives}}{\text{number of true negatives} + \text{number of false positives}}}$$

⇒ Since we have 2 classes

$$\text{Recall}_P = \frac{TP}{\text{Actual } P} \qquad \text{Recall}_N = \frac{TN}{\text{Actual } N}$$

$$\text{Precision}_P = \frac{TP}{\text{Predicted } P} \qquad \text{Precision}_N = \frac{TN}{\text{Predicted } N}.$$

$$\boxed{\text{macro Recall}} = \frac{1}{2}\left(\text{Recall}_P + \text{Recall}_N\right)$$

$$\boxed{\text{macro Precision}} = \frac{1}{2}\left(\text{Precision}_P + \text{Precision}_N\right)$$

$$\text{F1 Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

[NAT]

#Q. In the table below, the xi column shows scores on the aptitude test. Similarly, the yi column shows statistics grades. The last two columns show deviations scores - the difference between the student's score and the average score on each measurement. The last two rows show sums and mean scores.

H.W

Find the regression equation

$(.64x+27.08)$

Student	x_i	y_i	$(x_i-\bar{x})^2$	$(y_i-\bar{y})^2$
1	95	85	289	64
2	85	95	49	324
3	80	70	4	49
4	70	65	64	144
5	60	70	324	49
Sum	390	385	730	630
Mean	78	77		

[NAT]

#Q. The values of y and their corresponding values of y are shown in the table below.

X	0	1	2	3	4
y	2	3	5	4	6

H.W

(a) Find the least square regression line $y = a x + b$.

(b) Estimate the value of y when $x = 10$.

[MCQ]

#Q. Consider the linear regression model $Y = X\beta + \varepsilon$ with $\varepsilon \sim N(0n, \sigma\varepsilon^2\ Inn)$. This model (without intercept) is fitted to data using the ridge regression estimator $\hat{\beta}(\lambda) = \arg\min\beta \|Y - X\beta\| 2\ 2 + \lambda\|\beta\| 2\ 2$ with $\lambda > 0$.

The data are:

$$X^T = (-1\ 1\ 1\ -1) \text{ and } Y^T = (-1.5\ 2.9\ -3.5\ 0.7)$$

What is the maximum likelihood/ordinary least squares estimator of the regression parameter for $\lambda = 0$?

H.W

A [-0.3, 0.05]

B [-0.5, 0.1]

C [0.1, -0.2]

D [0.05, -0.3]

[MCQ]

#Q. A logistic regression model was used to assess the association between CVD and obesity. P is defined to be the probability that the people have CVD, obesity was coded as 0=non obese, 1=obese. log(P/(1-P)) = -2 + 0.7(obesity) What is the log odds ratio for CVD in persons who are obese as compared to not obese? (one correct choice)

Sol •

CVD and obesity → 0/1

$\log\left(\frac{P}{1-P}\right) = -2 + .7\,(\text{obesity})$

log odd for CVD : obese ⇒ $-2 + .7 = -1.3$

" " " " : non " ⇒ $-2 + .7(0) = -2$

$$\frac{\text{log odd Ratio for obese}}{\text{not obese}}$$

[MCQ]

#Q. A logistic regression model was used to assess the association between CVD and obesity. P is defined to be the probability that the people have CVD, obesity was coded as 0=non obese, 1=obese. log(P/(1-P)) = -2 + 0.7(obesity) What is the log odds ratio for CVD in persons who are obese as compared to not obese? (one correct choice)

Sol.

CVD and obesity → 0/1

P: Prob of CVD

log odd $\cdot \log \frac{P}{1-P} = -2 + .7(\text{Obesity})$

logodd Ratio for obese / not obese

logodd : obese ⇒ $-2 + .7(1) \Rightarrow -1.3$

logodd : nonobese ⇒ $-2 + .7(0) \Rightarrow -2$

Ratio ⇒ $-1.3/-2 \Rightarrow 1.3/2$

[MCQ]

#Q. Which of the following formula produces the correct value for the probability of having CVD (Cardiovascular Disease) from the logistic regression equation $\boxed{\log(P/(1-P)) = -2 + 0.7(\text{obesity})}$, $P_{CVD} \Rightarrow$

A. ~~Pcvd= exp(-2+.7o)/1- exp(-(-2+.7o))~~

B. $\boxed{\text{Pcvd= exp(-2+.7o)/ 1+ exp(-2+.7o)}}$ ✓

C. Pcvd= exp(-2+.7o)/ 1+ exp(-(-2 x .7o))

D. Pcvd= exp(-2 x.7o)/ 1+ exp(-(-2+.7o))

E. Pcvd= exp(-2+.7o)/ 1+ exp(-(-2+.7o))

$$P \Rightarrow \frac{1}{1+e^{-(\beta_0+\beta_1 x)}}$$

$$\boxed{\log_e \frac{P}{1-P} = \beta_0+\beta_1 x}$$

$$P = \frac{1}{1+e^{-(-2+.7\times \text{obesity})}}$$

$$P = \frac{e^{-2+.7\times \text{obe}}}{1+e^{-2+.7\times \text{obe}}}$$

Topic : Machine Learning

#Q. Which of the following may help to reduce $\boxed{\text{underfitting}}$ demonstrated by a model.

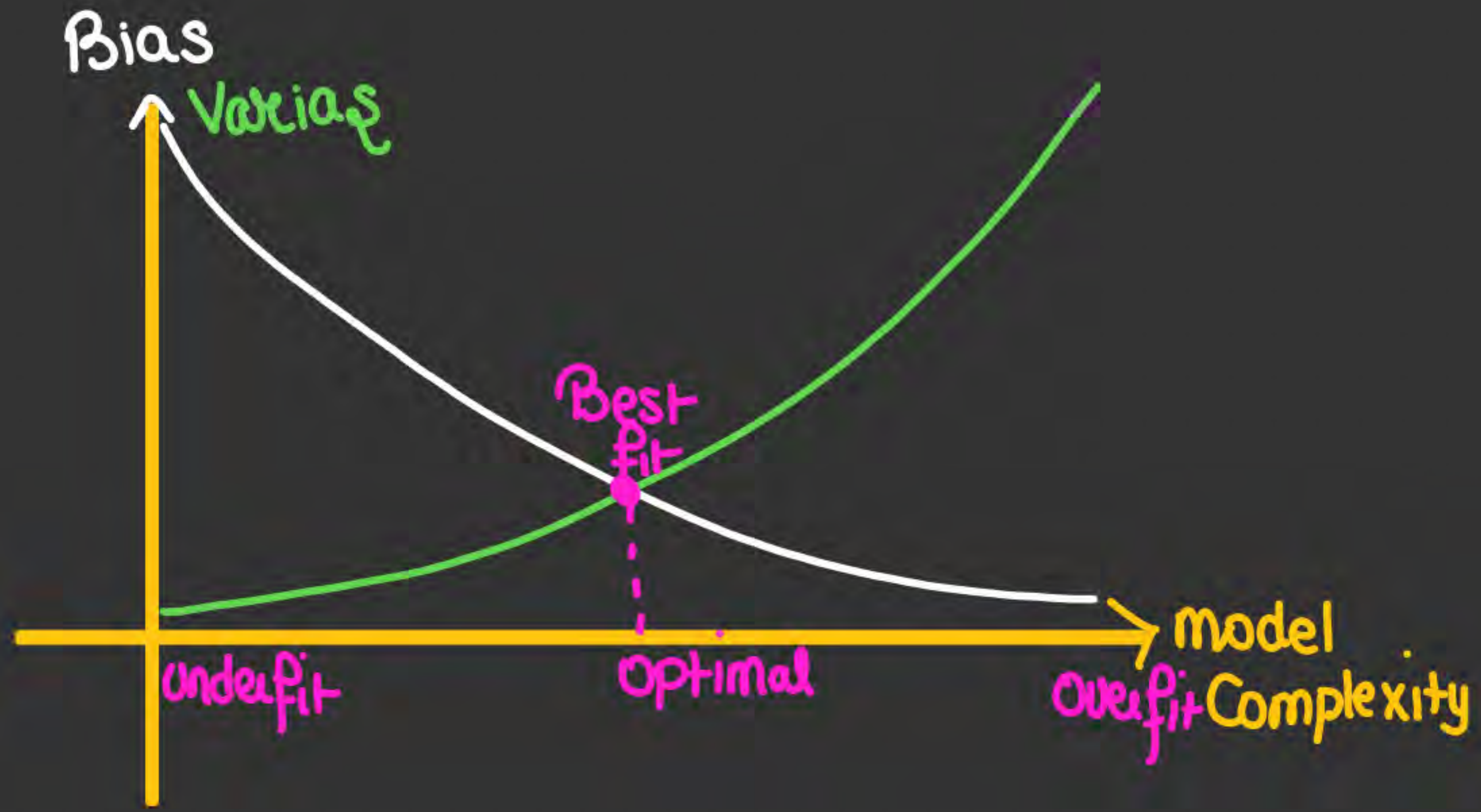
- S_1: Increase model complexity.
- S_2: Increase the number of optimization routine step.

A Only S_1

B Both S_1 and S_2 ✓

C Only S_2

D Neither S_1 nor S_2


Bias
Variace
Best fit
model Complexity
underfit
optimal
overfit

Topic : Machine Learning

#Q. Which of the following statements regarding bias and variance are true?

lowbias

A. Models which overfit have a high bias and underfit have a high variance.

B. Models which overfit have a high bias and underfit have a low variance

C. Models which overfit have a low bias and underfit have a high variance.

D. Models which overfit have a low bias and underfit have a low variance. ✓

Topic : Machine Learning

$P = \frac{TP}{\text{Total Predicted P}}$, $R = \frac{TP}{\text{Total Actual P}}$.

#Q. Suppose I have $\boxed{\text{10,000 emails}}$ in my mailbox out of which $\boxed{\text{200 are spams.}}$ The spam detection system detects $\boxed{\text{150 emails as spams,}}$ out of which 50 $\boxed{\text{are actually spam.}}$ What is the precision and recall of my spam detection system?

$P = \frac{TP}{TP + FN} = \frac{50}{100+50} = 1/3$; $R = \frac{TP}{TP+FN} = \frac{50}{200} = 1/4$

		Actual P	Actual N
Pred	P	50	100
	N	150	9700

A. $\boxed{\text{Precision 33.333\%, Recall = 25\%}}$ ✓

B. Precision = 25%, Recall = 33.33%

C. Precision = 33.33%, Recall = 75%

D. Precision = 75%, Recall = 33.33%

Topic : Machine Learning

#Q. Consider a binary classification task where you are evaluating a model's performance on a dataset. Let's denote the following:

- True Positives (TP): 12
- False Positives (FP): 5
- True Negatives (TN): 40
- False Negatives (FN): 8

Calculate F-score for the above data

Pred \ Actual	P	N
P	TP 12	FP 5
N	FN 8	TN 40

$$P = \frac{12}{12+5}, \quad R = \frac{12}{12+8} \quad R = .6$$

$$P = .705$$

$$F_{score} \rightarrow 2\frac{PR}{P+R} \rightarrow .648$$

Topic : Machine Learning

#Q. Which of the following statements is/are true?

B, D ✓

A ✗ Regularization in Ridge Regression involves adding a penalty proportional to the sum of the absolute values of the coefficients.

square

B ✓✓ Principal Component Analysis (PCA) reduces dimensionality by transforming features into a new set of orthogonal components.

C ✗ Random Forests use a single decision tree to make predictions.

D ✓ Gradient Descent is used to optimize the parameters in a neural network.

Topic : Machine Learning

#Q. Given the following information

	Actual Positive	Actual Negative
Predicted Positive	15	35
Predicted Negative	10	45

What is the precision and recall?

$P = \frac{15}{15+35} = .3$, $R = \frac{15}{15+10}$ $.6$

A. 0.5, 0.6

B. 0.3, 0.8

C. 0.3, 0.6

D. 0.5, 0.8

#Q. For a binary classification problem, consider the two statements below:

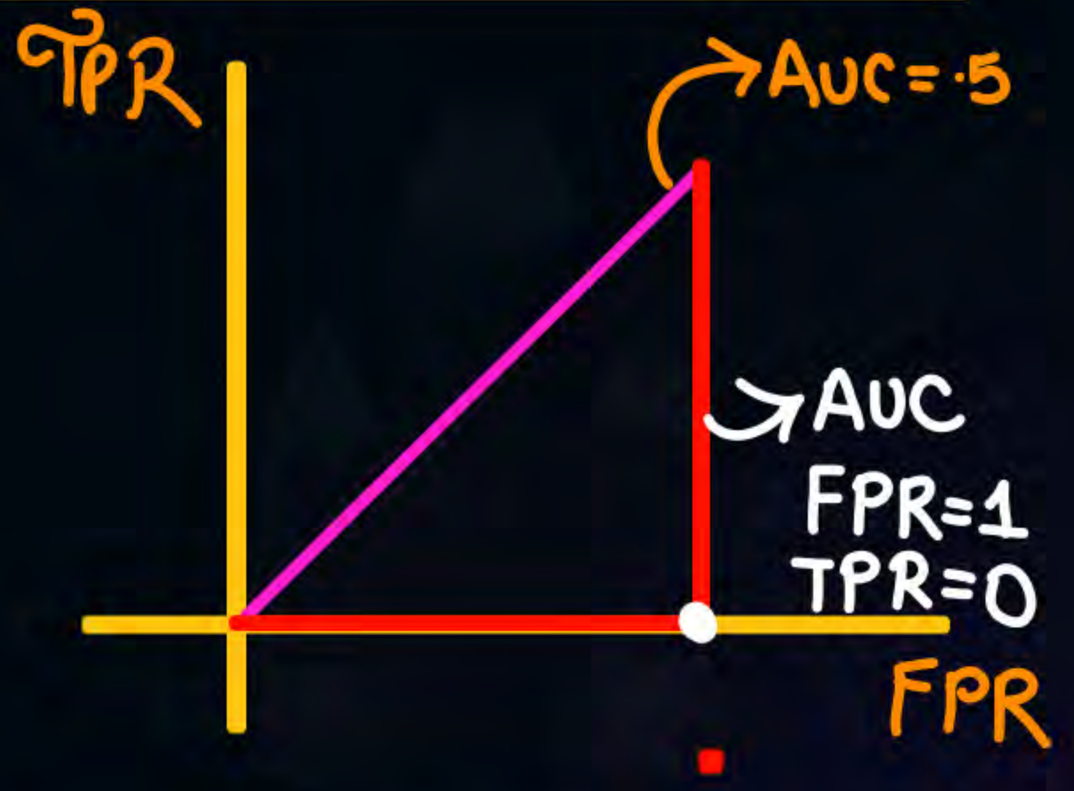
A: A classifier with AUC=0 is the least useful classifier. → Can work as Perfect classifier

~~B~~: A classifier with AUC=0.5 is the least useful classifier

→ Random Prediction

[redacted]

D) Both are true



A A is True. B is False

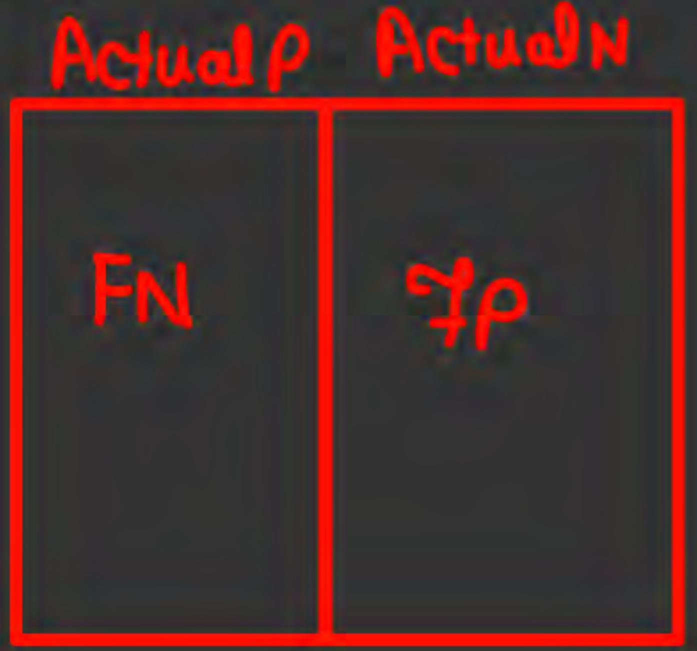
B ✓ A is False. B is True

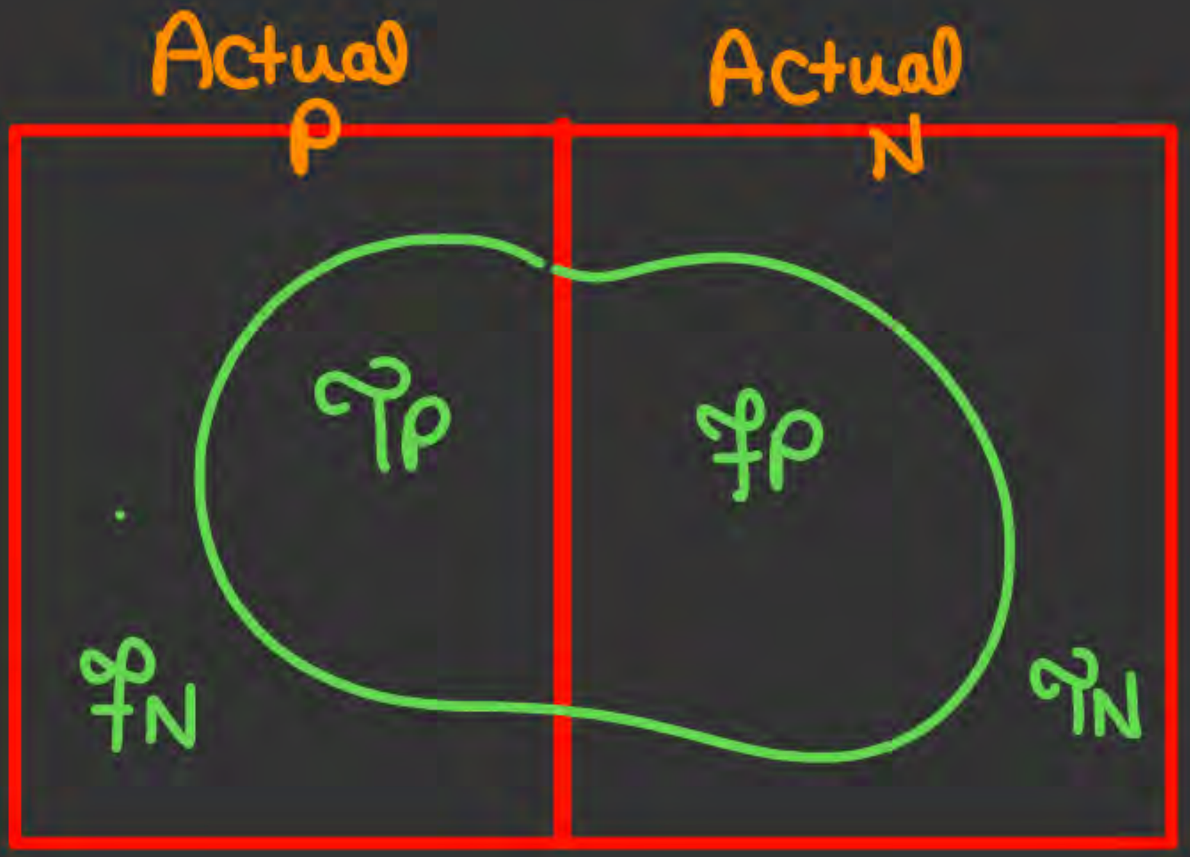
C Both are False

[redacted]

$$TPR = \frac{TP}{\text{Actual } P} = 0$$

$$FPR = \frac{FP}{\text{Actual } N} = 1$$

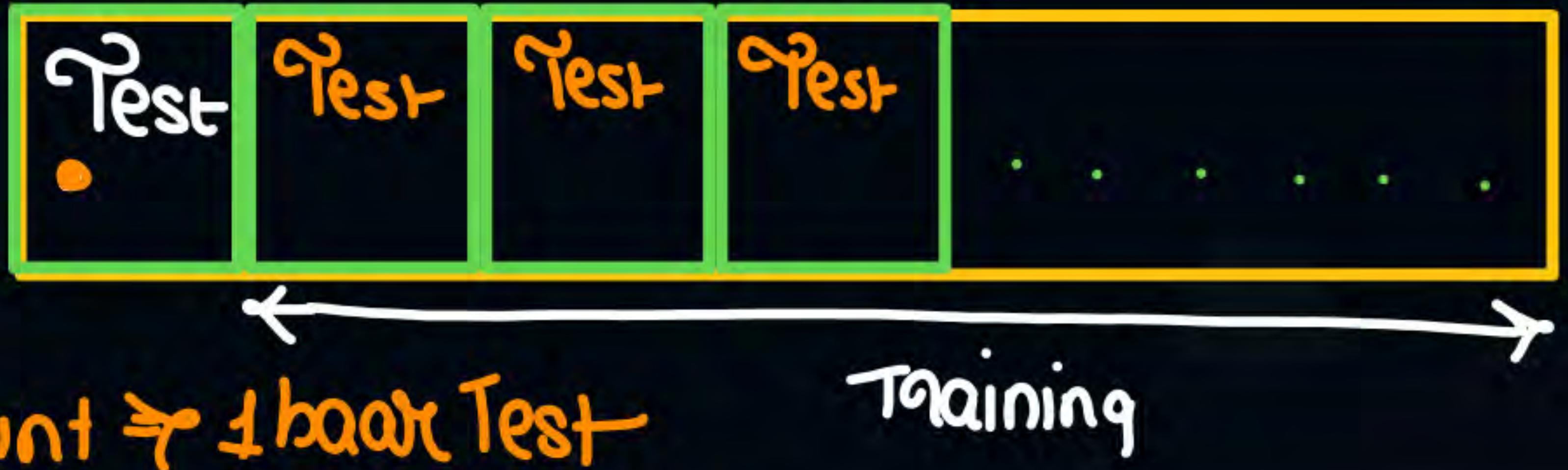




Topic : Machine Learning

#Q. If you use 10 - fold cross - validation on a dataset of 1,000 samples, how many times will each sample be used for testing?

Break data into 10 Parts



each data point → 1 baar Test
9 baar training

- **A** 10 times
- **B** 1 time
- **C** 5 times
- **D** 100 times

Topic : Classification

#Q. Imagine you have a dataset that contains information about a binary classification model. The model classifies instances into two classes: "Positive" and "Negative". After evaluating the model, you have the following results:

H.W

True Positives (TP): 120

False Positives (FP): 25

True Negatives (TN): 800

False Negatives (FN): 30

Calculate the precision of the model using the given information.

A 0.75

B 0.80

C 0.82

D None of the above

Topic : Classification

#Q. Which of the following is false about a logistic regression based classifier?

logit fxn
(logodds ⇒ $\beta_0+\beta_1 x$)

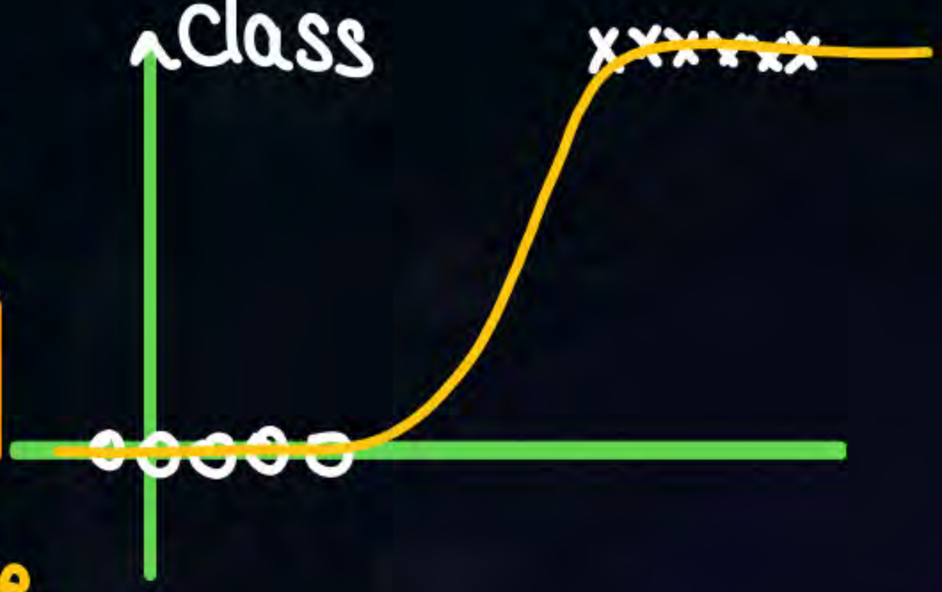
False

A. The logistic function is non-linear in the weights

B. The logistic function is linear in the weights — True

C. The decision boundary is non-linear in the weights — True

D. The decision boundary is linear in the weights — false.



Topic : Classification

#Q. In continuation with question 7, let x = 1 if the server is wearing black shirt and x = 0 for servers wearing other colored shirts. We know that there are 70 observations with x 1 and 340 observations with x = 0. The response variable is also an indicator variable given by y= 1 if the customer left a tip and y=0 if the customer did not leave a tip. Use this data to fit a logistic regression model to compute the log- odds of leaving a tip depending on the color of the server's shirt, when server wear black shirt Cases of $y=1 \Rightarrow 45$, and when server wear other shirt Case of $y=1 \Rightarrow 150$.

$\Rightarrow$ $\log odd = \beta_0 + \beta_1 x$

$x=1: \; P_y = 45/70$ ✓

$x=0: \; P_y = 150/340$

- Independent Var $\Rightarrow$ Shirt Color (x)
- dependent Var $\Rightarrow$ Tip (y)

$$\log \frac{P}{1-P} = \beta_0 + \beta_1 x$$

eq 1 $x=0$

$$\log \frac{P_y}{1-P_y} = \beta_0$$

$$-.236 = \beta_0$$

eq 2 $x=1$

$$\log \frac{P}{1-P} = \beta_0 + \beta_1$$

$$.5877 = -.236 + \beta_1$$

$$\beta_1 = .823$$ ✓

#Q. For a linear classifier you have the following weights and bias:

$w = [2, -3]$ and $b = 1$.

What is the classification result for the input vector $x = [1,1]$?

Classifier $(2x^1 - 3x^2 + 1)$

$(1,1) \Rightarrow 2(1) - 3(1) + 1 \Rightarrow 0$

$\omega^T x + B > 0$ Class 1

$\omega^T x + B < 0$ Class 0

$\omega^T x + B = 0 \rightarrow$ Point is on the Classifier

A. Class 0

B. Class 1

C. Class 2

D. $\boxed{\text{It cannot be classified}}$ ✓

Topic : Classification

#Q. A linear classifier has the following weight vector w =[1, –2, 3] and bias b = 4. What is the value of the decision function for the point x = [2, –1, 3]?

$w_1x^1 + w_2x^2 + w_3x^3 + b \Rightarrow$ value

$2\times1 - 1\times-2 + 3\times3 + 4$

$2+2+9+4 \Rightarrow 17$

Topic : Classification

#Q. Given the equation of a linear classifier: $3x - 4x + 2 = 0$.

What is the output of the classifier for the input vector $(x, y) = (1,2)$?

H.W

A. Class 0

B. Class 1

C. The classifier is undefined

D. The point on the boundary

#Q. Accuracy is simply a ratio of correctly predicted observations to the total observations. From the above confusion matrix, how would you define Accuracy?

H.W

A) Accuracy = (FP+FN)/(TP+FN+FP+TN)

B) Accuracy = (TP+TN)/(TP+FN+FP+TN)

C) Accuracy = (TP+FN)/(TP+FN+FP+TN)

D) Accuracy = (FP+TN)/(TP+FN+FP+TN)

#Q. In Ridge Regression, what is the effect of $\boxed{\text{increasing } \lambda \text{ (lambda)}}$ on the bias and variance, of the model?

A. Increases bias, decreases variance.

B. Decreases bias, increases variance.

C. Increases both bias and variance.

D. $\boxed{\text{Decreases both bias and variance.}}$

$\lambda = 0$

overfit → Best fit

if λ is v. large

underfit

#Q. Ridge Regression can help prevent overfitting, but what is the trade-off?

A. Increased model interpretability

B. $\boxed{\text{Increased computational complexity}}$

C. Reduced accuracy on the training data

D. Smaller training dataset size

#Q. Consider the data collected from 410 customers in a restaurant. It is observed that 40 of the 70 customers tipped the server who was wearing a black shirt and 130 of the 340 customers tipped the server who was wearing a different color. Compute the logit or log-odds of tipping a server wearing a black shirt.

H.W

A 0.2877

B 0.1249

C –0.7677

D –1.7677

#Q. In continuation with question 7, let x = 1 if the server is wearing black shirt and x = 0 for servers wearing other colored shirts. We know that there are 2 70 observations with x = 1 and 340 observations with x = 0 The response variable is also an indicator variable given by y = 1 if the customer left a tip and y = 0 if the customer did not leave a tip. Use this data to fit a logistic regression model to compute the log-odds of leaving a tip depending on the color of the server's shirt..

P.W

A. $-0.4797 + 0.1249x$

B. $0.2877 + 0.1249x$

C. $0.1249 + 0.4317x$

D. $-0.4797 + 0.7674x$

#Q. You are given a trained Logestic Regression model with the following numerical weight vector and bias:

weight vector (w): [0.8, –1.2]

Bias (b): -0.5

You need to classify four points (A, B, C, D) using this model. The data points and their respective feature vectors are as follows:

Point A: (3, 5] $\rightarrow -4.1$ Valid $\boxed{\begin{matrix} w^T x + b > 0 \rightarrow \text{Class } 1 \\ w^T x + b < 0 \rightarrow \text{Class } 0 \end{matrix}}$ $P_1 = \frac{1}{1+e^{-(w^T x + b)}}$

Point B: (–2, 4] $\rightarrow -6.9$ $P_1 > .5 = \text{Class } 1$

Point C: [1, –1] $\rightarrow 1.5$ Class 1. $P_1 < .5 = \text{Class } 0$

Point D: [–4, –3] $\rightarrow -.1$

Which points will be classified as Class 1 (positive class) using this Logistic Regression model?

$$\text{MLE} \Rightarrow \max \prod_{i=1}^{N} (P_i)^{y_i} (1-P_i)^{(1-y_i)} \Rightarrow$$

max(Product of Probab of Class 1 for class 1 points
X
Probab of Class 0 for Class 0 points)

P_i = Probab of Class 1 for i^{th} Point

Class 1 point $y_i = 1$

Class 0 point $y_i = 0$

#Q. The following table gives the binary labels ($y^{(i)}$) for four points $\left(x_1^{(i)}, x_2^{(i)}\right)$ where $i = 1, 2, 3, 4$. Among the given options, which set of parameter values $\beta_0, \beta_1, \beta_2$ of a standard logistic regression model $p(x_i) = \frac{1}{1+e^{-(\beta_0+\beta_1 x+\beta_2 x)}}$ results in the highest likelihood for this data?

A $\beta_0 = 0.5, \beta_1 = 1.0, \beta_2 = 2.0$ ×

$P_1 P_1 P_0 P_0 \Rightarrow$

$\frac{1}{1+e^{-(.5)}} \quad \frac{1}{1+e^{-.1}} \quad \left(1 - \frac{1}{1+e^{-1.8}}\right) \quad \left(1 - \frac{1}{1+e^{-.8}}\right)$

$\Rightarrow .0143$

	x_1	x_2	y
+ve	0.4	-0.2	1
+ve	0.6	-0.5	1
+ve	-0.3	0.8	0
+ve	-0.7	0.5	0

#Q. The following table gives the binary labels ($y^{(i)}$) for four points $\left(x_1^{(i)}, x_2^{(i)}\right)$ where $i = 1, 2, 3, 4$. Among the given options, which set of parameter values $\beta_0, \beta_1, \beta_2$ of a standard logistic regression model $p(x_i) = \frac{1}{1+e^{-(\beta_0+\beta_1 x+\beta_2 x)}}$ results in the highest likelihood for this data?

A $\beta_0 = 0.5, \beta_1 = 1.0, \beta_2 = 2.0$

B $\beta_0 = -0.5, \beta_1 = -1.0, \beta_2 = 2.0$

[redacted]

	x_1	x_2	y
-ve	0.4	-0.2	1
-ve	0.6	-0.5	1
+ve	-0.3	0.8	0
+ve	-0.7	0.5	0

#Q. The following table gives the binary labels ($y^{(i)}$) for four points $\left(x_1^{(i)}, x_2^{(i)}\right)$ where i = 1, 2, 3, 4. Among the given options, which set of parameter values $\beta_0, \beta_1, \beta_2$ of a standard logistic regression model $p(x_i) = \frac{1}{1+e^{-(\beta_0+\beta_1 x+\beta_2 x)}}$ results in the highest likelihood for this data?

A $\beta_0 = 0.5, \beta_1 = 1.0, \beta_2 = 2.0$

B $\beta_0 = -0.5, \beta_1 = -1.0, \beta_2 = 2.0$

C $\beta_0 = 0.5, \beta_1 = 1.0, \beta_2 = -2.0$

D [redacted]

	x_1	x_2	y
+ve 1.3	0.4	–0.2	1
2.1	0.6	-0.5	1
-1.4	–0.3	0.8	0
-1.2	–0.7	0.5	0

#Q. The following table gives the binary labels ($y^{(i)}$) for four points $\left(x_1^{(i)}, x_2^{(i)}\right)$ where i = 1, 2, 3, 4. Among the given options, which set of parameter values $\beta_0, \beta_1, \beta_2$ of a standard logistic regression model $p(x_i) = \frac{1}{1+e^{-(\beta_0+\beta_1 x+\beta_2 x)}}$ results in the highest likelihood for this data?

A $\boxed{\beta_0 = 0.5, \beta_1 = 1.0, \beta_2 = 2.0}$

B $\beta_0 = -0.5, \beta_1 = -1.0, \beta_2 = 2.0$

C $\beta_0 = 0.5, \beta_1 = 1.0, \beta_2 = -2.0$

D $\boxed{\beta_0 = -0.5, \beta_1 = 1.0, \beta_2 = 2.0}$

	x_1	x_2	y
-ve ←	0.4	-0.2	1
-ve ←	0.6	-0.5	1
+ve ←	-0.3	0.8	0
-ve ←	-0.7	0.5	0

#Q. In logistic regression, what is the role of the logistic function $\boxed{\text{(sigmoid function)}}$?

A It transforms the independent variables.

B It models the relationship between the dependent and independent variables.

C ✓ It converts the log-odds into probabilities.

D It calculates the likelihood of the data. ✗

log odd ⇒ $\beta_0 + \beta_1 x$

and Sigmoid Convert this $\beta_0 + \beta_1 x$ into Probab

$$P = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

#Q.

n = 200	Prediction = NO	Prediction = YES
Actual = NO	60	10
Actual = YES	5	125

Yes: Balance > 1000 \$, No: Balance < 1000 \$

A In reality, there are totally 135 accounts who have a balance more than \$1000 and 70 accounts with balance - less than \$1000

B In reality, there are totally 60 accounts who have a balance more than \$1000 and 70 accounts with balance - less than \$1000

C In reality, there are totally 125 accounts who have a balance more than \$1000 and 10 accounts with balance - less than \$1000

D In reality, there are totally 130 accounts who have a balance more than \$1000 and 70 accounts with balance - less than \$1000

#Q. For the below confusion matrix, what is the recall?

	NOT 5	5
NOT 5	5578	1345
5	1234	3452

#Q. We will have 2 dimension

Apply Logistic regression and find the classifier

The data is as follows...		
x^1	x^2	Psuccess
1	5	.3
2	3	.2
5	6	.8

$$\boxed{\log_e odd = \beta_0 + \beta_1 x^1 + \beta_2 x^2}$$

$$\ln(.3/.7) = \beta_0 + \beta_1 + 5\beta_2$$

$$\ln(.2/.8) = \beta_0 + 2\beta_1 + 3\beta_2$$

$$\ln(.8/.2) = \beta_0 + 5\beta_1 + 6\beta_2$$

$\beta_0 = -3.72$

$\beta_1 = 0.436$

$\beta_2 = .4877$

log odd

$$\log \frac{P_s}{1-P_s} \Rightarrow \left(\log_e \frac{.3}{.7}\right)$$

H/W

#Q. For the below confusion matrix, what is the F_1 score?

	NOT 5	5
NOT 5	5578	1345
5	1234	3452

#Q. For a model to detect videos that are unsafe for kids, we need (safe video = postive class)

- [x] High precision, low recall
- [ ] High recall, low precision

FP is ≈ 0

FN $\neq 0$

$$P = \frac{TP}{TP+FP} \approx 1$$

$$R = \frac{TP}{TP+FN} \approx \text{Can less}$$

		Actual P	Actual N
Pred	P	TP	FP
	N	FN	TN

#Q. For the below confusion matrix, what is the precision?

	NOT 5	5
NOT 5	5578	1345
5	1234	3452

H.W

2 mins Summary

Topic Machine Learning

Topic

THANK - YOU