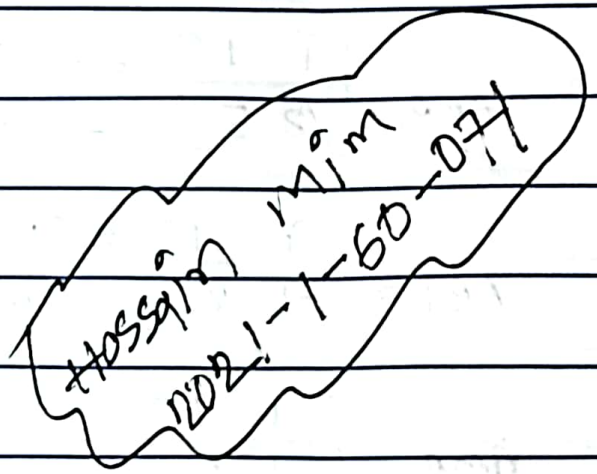
Final :—

1) confusing Matrix
2) Regression
3) Kernel (svm) বুঝতে
4) • matrix এর
বিয়োগ
যুক্ত
বিপরীত



• multiplication associative
• Not multiplication associative

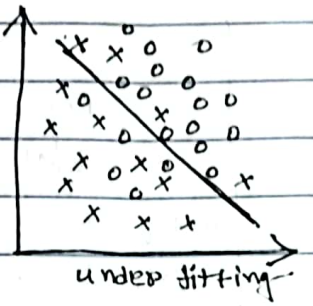
(প্রমাণ)

Dmim SIR OF BOB Final

Date :
Page :

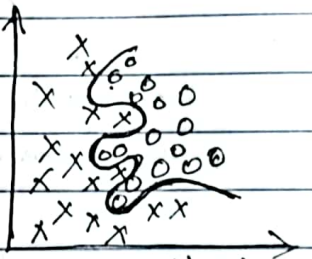
Under-fitting :- when a model has not learned learned the patterns in the training data well and is unable to generalization well on the new data, it is known as under-fitting. An under-fitting model has poor performance will result in unreliable. under-fitting occurs due to high bias and low variance. It too simple model and too much regulagation in the model.

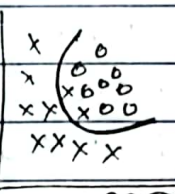
Improtant



Hossain Mim
2021-1-60-071

over-fitting :- over-fitting is a phenomenon that occurs when a machine learning model is constria constraint to training set and not able to perform well on unseen data. over-fitting model is very complex and too littel regutaz regularization over-fitting model occurs due to low bias and high variance





Bias :- we can ~~and~~ define bias the error between average model prediction and ground truth.

- A model which High-Bias would not match the data set closely
- A model Low-Bias will closely match the training data set

Date :
Page :

variance refers to the changes in the model when using different portions of the traning data set. It is the variability in the model prediction - how much the machine learning function can adjust - develops depending on the given data set.

Important table

Algorith	Bias	variance
• Linear regression	High	Less
• Decision tree	Low	High
• Bagging	Low	High (Less than Decision tree)
• Random forest	Low	High (Less than DT and B

Get more then data :—
- when we are have high variance

Important
Try different features :—
- Adding features helps fix High Bias
- using smaller set of features fix high variance

Try tuning your hyperparameter :—
- Decrease regularization when bias is high
- Increase regularization when variance is high.

when a model has not learned the patterns in the training data well and is unable to generalization well on the new data, it is known as under-fitting.

when a model has not learned the patterns in the training data well and is unable to generalization well on the new data.

Over-fitting is a phenomenon that occurs when a machine learning mode is constraint to training set and not able to perform well and is unable to generalization well on the new data.

over-fitting is a phenmenon that occurs when a ML model is constraint to training set and not able to perform well on unseen data.

Date :
Page :

• In multiplication is associative justify :-

$$x = \begin{bmatrix} A & B \\ C & D \end{bmatrix}_{2}^{1} \qquad y = \begin{bmatrix} E & F \\ G & H \end{bmatrix}_{1}^{2} \qquad z = \begin{bmatrix} I & J \\ K & L \end{bmatrix}$$

$$\{(x \times y) z\} = \{x \times (y \times z)\}$$

Now,

$$(x \times y) = \begin{bmatrix} AE + BG & AF + BH \\ CE + DG & CF + DH \end{bmatrix}$$

$$\{(x \times y) z\} = \begin{bmatrix} AE + BG & AF + BH \\ CE + DG & CF + DH \end{bmatrix} \times \begin{bmatrix} I & J \\ K & L \end{bmatrix}$$

$$= \begin{bmatrix} AEI + BGI + AFK + BHK & AEJ + BGJ + AFL + BHL \\ CEI + DGI + CFK + DHK & CEJ + DGJ + CFL + DHL \end{bmatrix}$$

$$(y \times z) = \begin{bmatrix} EI + FK & EJ + FL \\ GI + HK & GJ + HL \end{bmatrix}$$

$$\{x \times (yz)\} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \times \begin{bmatrix} EI + FK & EJ + FL \\ GI + HK & GJ + HL \end{bmatrix}$$

$$= \begin{bmatrix} AEI + AFK + BGI + BHK & AEJ + AFL + BGJ + BHL \\ CEI + CFK + DGI + DHK & CEJ + CFL + DGJ + DHL \end{bmatrix}$$

So, $\{x \times (yz)\} = \{(x \times y) \times z\}$ Ans

Matrix multiplication is associative.

Ans

Final ans (2)

Date :
Page :

In not commutative justify

$$x = \begin{bmatrix} a & b & p \\ c & d & q \end{bmatrix} \qquad y = \begin{bmatrix} e & f \\ g & h \\ i & j \end{bmatrix}$$

$(xy) \neq (yx)$

$$(xy) = \begin{bmatrix} a & b & p \\ c & d & q \end{bmatrix} \times \begin{bmatrix} e & f \\ g & h \\ i & j \end{bmatrix}$$

$$\begin{bmatrix} ae+bg+pi & af+bh+pj \\ ce+dg+qi & cf+dh+qj \end{bmatrix}$$

$$(yx) = \begin{bmatrix} e & f \\ g & h \\ i & j \end{bmatrix} \times \begin{bmatrix} a & b & p \\ c & d & q \end{bmatrix}$$

$$\begin{bmatrix} ea+fc & eb+fd & ep+eq \\ ga+hc & gb+hd & gp+hq \\ ia+jc & ib+jd & ip+jq \end{bmatrix}$$

so, $(xy) \neq yx$

we can say that the martix multipulction, is not commutative (Ans)

Final

Date :
Page :

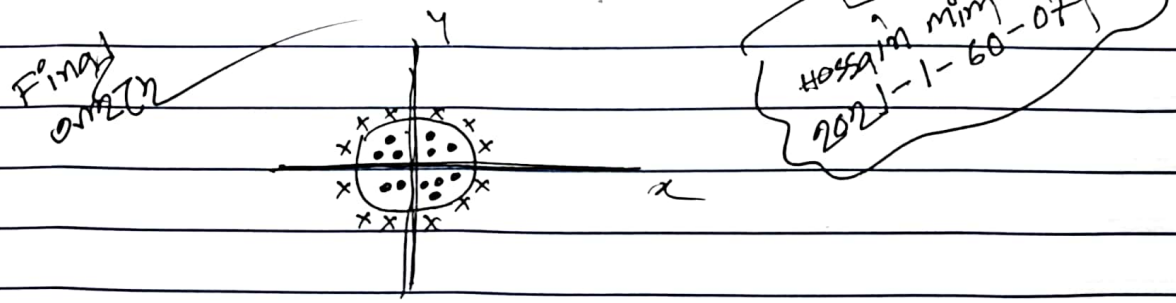
The support vector machine kernel is a function that takes low dimensional input space and transform it into a high dimensional space that means it converts not separable problem to a separable problem. It is mostly useful in non-linear data separation. It simply put it dose some extremely complex data transmation then finds out the process to separate data base on the labels or output

The svm kernel is a function that takes low dimensional input space and transform it into a high dimensional space that means it converts not separable problem to a separable problem. It is mostly useful in non-linear data separation problem. simply put it dose some extremely complex data transmation then finds out the process to separate the data based on the labels or output.

It works by finding the best possible boundaries that can separate two data points with maximum margin, also known is as hyperplane.

The svm kernal is a function that takes low dimensional input space and transform it into a high dimensional space that means it converts not separable problem to a separable problem. It is mostly useful in non-linear data separation problem. It simply put dose some extremely complex data transmation, then it finds out the process the data based on the labels or output.

It works by finding the best possible boundaries than can separate two classes data points with maximum margin, also it known as hyperplane.



Hossain Mim
202-1-60-07

Ans

SVM

Date :
Page :

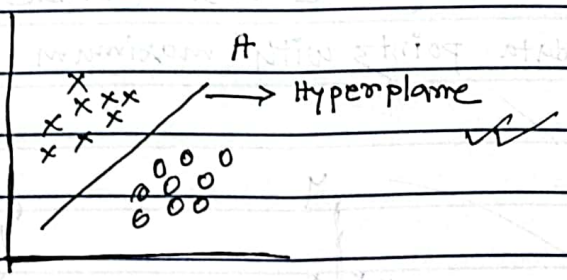
SVM is defined as a machine learning algorithms that uses supervised learning model to solve complex classification, regression, and outlier problem detection problem by performance optimal data transmations that determine boundaries between data point based on the labels or output.

~~The support vector machine is~~

The svm is defined as a machine learning algorithm that uses supervised learning model to solved complex classification regression and outlier detection problems by performing optimal data tansmation that determine boundaries between data point based on the labels or output.

support vector machine is defined as a machine learning that uses supervised learning model to solved complex classification, regression and outlier detection problem by performing ~~that determine boundaries~~ optimal data that determine boundaries between data point based on the output

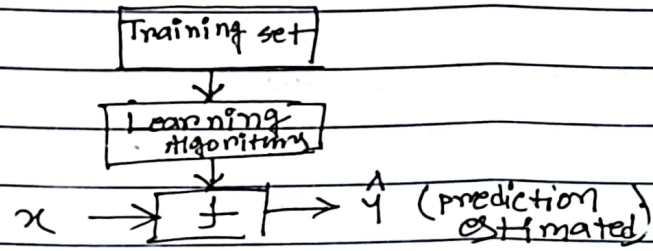
The support vector machine is defined as a machine learning algorithms that uses supervised learning model to solved complex classification, regression and outlier detection problem by performing the optimal data transmation that determine boundaries between the data points based on the labels or output



7

Date :
Page :

Traing set: Traing set is the major portion of orginal dataset which is used for ~~testing~~ training the ML model.



Important

- model works accuracy $y = \hat{y}$
- gap between y and $\hat{y}$ is known as error

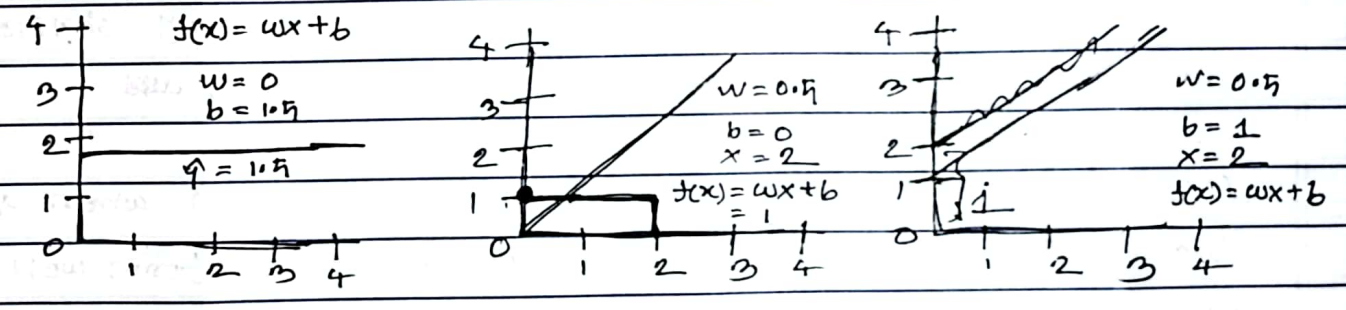
Equation $= f_{w,b}(x) = wx + b$

$= f(x) = wx + b$

$= \hat{y}$

cost function: Cost function quantifies the error between predicted and target values and present that error in the from of a single real number

Imp

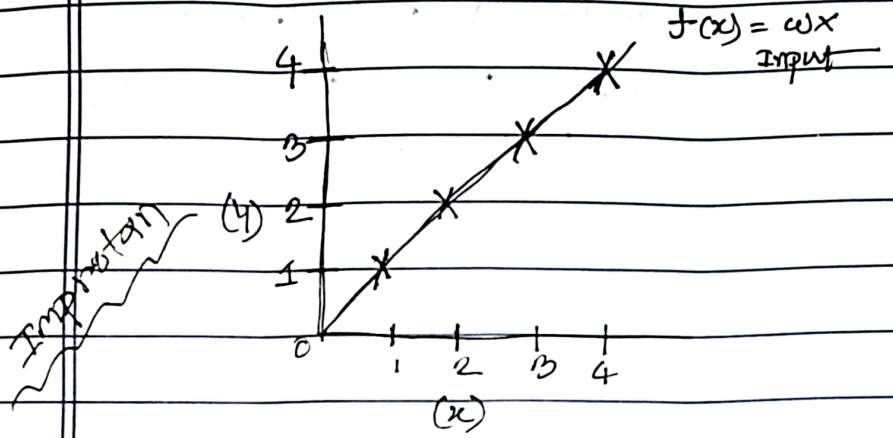


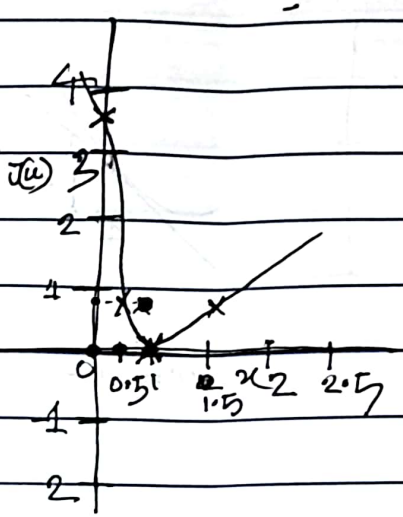
cost function:

$$J(w,b) = \frac{1}{2m} \sum_{i=1}^{m} (\hat{y}^{(i)} - y^{(i)})^2$$

m = number of training example.

Regression problem :—





when, $w = 0$

$f(x) = wx$

$= (0 \times 1) = (0 \times 2) = (0 \times 3) = (0 \times 4)$

$\Rightarrow 0 \quad \Rightarrow 0 \quad \Rightarrow 0 \quad = 0$

Now, $J(w) = \frac{1}{2 \times 4} \{ (0-1)^2 + (0-2)^2 + (0-3)^2 + (0-4)^2 \}$

$= 3.75$

when $w = 0.5$

$\Rightarrow (0.5 \times 1) = (0.5 \times 2) \Rightarrow (0.5 \times 3) = (0.5 \times 4)$

$= 95 \quad \Rightarrow 1.0 \quad = 1.5 \quad = 2.0$

Now, $J(w) = \frac{1}{2 \times 4} \{ (0.5-1)^2 + (1.0-2)^2 + (1.5-3)^2 + (2-4)^2 \}$

$= 0.94$

when, $w = 1$

$\Rightarrow (1 \times 1) \Rightarrow (1 \times 2) \Rightarrow (1 \times 3) \Rightarrow (1 \times 4)$

$= 1 \quad \Rightarrow 2 \quad \Rightarrow 3 \quad = 4$

$\Rightarrow J(w) = \frac{1}{4 \times 2} \{ (1-1)^2 + (2-2)^2 + (3-3)^2 + (4-4)^2 \}$

$= 0$

when $w = 1.5$

$= 1.5 \times 1 \Rightarrow 1.5 \times 2$

$= 1.5 \quad = 3$

$= 1.5 \times 3 = 1.5 \times 4$

$= 4.5 \quad = 6$

$J(w) = \frac{1}{2 \times 4} \{ (1.5-1)^2 + (3-2)^2 + (4.5-3)^2 + (6-4)^2 \}$

$\Rightarrow 0.94$

So, we can see lowest w value is $w = 1$

Best regression line is $w = 1$ (Ans)

GD

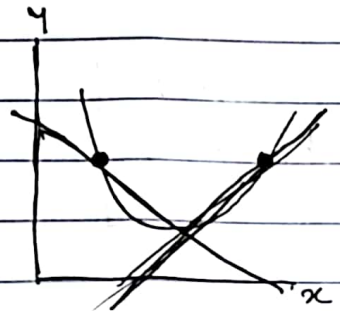
Date :
Page :

GD Algo Problem :—

we know that

$$w = w - \alpha \frac{\partial}{\partial w} J(w,b)$$

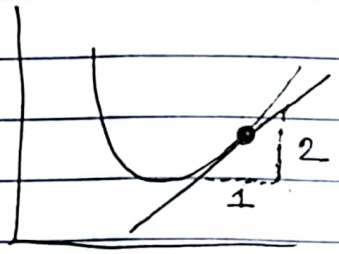
weight — learing rate



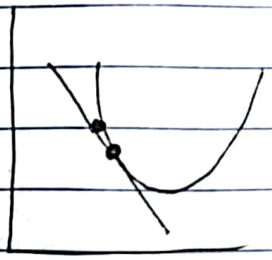
Assum,
$w = 2$
$\alpha = -1$
$w = \{2 - (-1)\}$
$= 2+1$
$= 3$

Assum,
$w = 2$
$\alpha = 1$
$w = (2-1)$
$= 1$

So, we can say that the learnig rate (α) is negative thats for weight value is increass And (α) learning positive(-) thats for (w) weight is decrease (Ans)



$w = w - \alpha$ (positive number)



$w = w - \alpha$ (negative number)

Hossain mim
2021-1-60-071

- If (α) learning rate is too small the GD is may be slow
- If (α) learning rate is too large The GD may be overshoot, never minimum and fail to converge

Date :
Page :

Inverse matrix:—

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 2 \\ 3 & 5 & -7 \end{bmatrix} \quad x = \begin{bmatrix} 8 \\ 6 \\ 14 \end{bmatrix}$$

$Ax = B$

$x = A^{-1}B$

$|A| \Rightarrow 1C$

	1	1	1
$\|A\|$	1	-1	2
	3	5	-7

Actual	predicted: Crocodile	Alligator
Crocodile (P)	440 (TP)	60 (FN)
Alligator (N)	80 (FP)	420 (TN)

For crocodile:—

1) Accuracy $= \frac{TP + TN}{TP + TN + FN + FP}$

$= \frac{440 + 420}{440 + 420 + 80 + 60}$

$= 0.86$

$= 86\%$ (Ans)

2) Recall $= \frac{TP}{TP + FN}$

$= \frac{440}{440 + 60}$

$= 0.88$

$= 88\%$ (Ans)

3) Precision $= \frac{TP}{TP + FP}$

$= \frac{440}{440 + 80}$

$= 0.846$

$= 84.6\%$ (Ans)

4) F1-score $= 2 \times \frac{0.846 \times 0.88}{0.846 + 0.88}$

$= 0.862$

$= 86.2\%$ (Ans)

For Alligator:—

1) Accuracy $= \frac{TN + FP}{TN + FP + FN + TP}$

$= \frac{420 + 80}{420 + 80 + 440 + 60}$

$= 0.5$

$= 56\%$ (Ans)

2) Recall $= \frac{TN}{TN + FP}$

$= \frac{420}{420 + 80}$

$= 0.84$

$= 84\%$ (Ans)

3) Precision $= \frac{TN}{TN + FN}$

$= \frac{420}{420 + 60}$

$= 0.875$

$= 87.5\%$ (Ans)

4) F1-score $= 2 \times \frac{0.875 \times 0.84}{0.875 + 0.84}$

$= 0.857$

$= 85.7\%$

(This suggest that the classifier perform better in identifying "x" instances than "y" instances, especially in terms of pres and recall.