1. **Bias:** Bias refers to the error introduced by approximating a real-world problem, which may be very complex, by a simpler model. It represents the difference between the average prediction of our model and the correct value which we are trying to predict.
    - **Formula:** Bias is not represented by a specific mathematical formula but is observed in the difference between predicted values and actual values. In mathematical terms, bias is often related to how well the model fits the training data.
2. **Variance:** Variance represents the variability of model predictions for a given data point or a range of data points. It measures the model's sensitivity to fluctuations in the training set.
    - **Formula:** Variance can be mathematically represented as the expected value of the squared difference between the predicted value and the average predicted value: $\mathrm{Var}(X) = E[(X - E[X])^2]$

Regularization: Regularization is a technique used to prevent overfitting by adding a penalty term to the model's loss function. It discourages the learning algorithm from fitting the model too closely to the training data and helps in generalizing the model.

- **L2 Regularization (Ridge):** Adds the square of the magnitude of coefficients to the loss function: $\mathrm{Loss} + \lambda \sum_{i=1}^{n} \beta_i^2$
- **L1 Regularization (Lasso):** Adds the absolute value of coefficients to the loss function: $\mathrm{Loss} + \lambda \sum_{i=1}^{n} |\beta_i|$

NAIVE BAYES CLASSIFIER - EXAMPLE - 2

$P(M) = \frac{4}{8} = 0.5 \quad P(H) = \frac{4}{8} = 0.5$

Color	M	H
White	2/4	3/4
Green	2/4	1/4

Legs	M	H
2	1/4	4/4
3	3/4	0/4

Height	M	H
Tall	3/4	2/4
Short	1/4	2/4

Smelly	M	H
Yes	3/4	1/4
No	1/4	3/4

$p(M|New\ Instance) = p(M) * \underline{p(Color = Green|M)} * p(\underline{Legs = 2}|M) * p(\underline{Height = tall}|M) * p(\underline{Smelly = no}\ |M)$

given

$p(M|New\ Instance) = 0.5 * \frac{2}{4} * \frac{1}{4} * \frac{3}{4} * \frac{1}{4} = 0.0117$

$p(H|New\ Instance) = p(H) * p(Color = Green|H) * p(Legs = 2|H) * p(Height = tall|H) * p(Smelly = no\ |H)$

$p(H|New\ Instance) = 0.5 * \frac{1}{4} * \frac{4}{4} * \frac{2}{4} * \frac{3}{4} = 0.047$

$\boldsymbol{p(H|New\ Instance) > p(M|New\ Instance)}$

Hence the new instance belongs to Speices H

KNN Classifier Solved Example - 1

Sepal Length	Sepal Width	Species
5.3	3.7	Setosa
5.1	3.8	Setosa
7.2	3.0	Virginica
5.4	3.4	Setosa
5.1	3.3	Setosa
5.4	3.9	Setosa
7.4	2.8	Virginica
6.1	2.8	Verscicolor
7.3	2.9	Virginica
6.0	2.7	Verscicolor
5.8	2.8	Virginica
6.3	2.3	Verscicolor
5.1	2.5	Verscicolor
6.3	2.5	Verscicolor
5.5	2.4	Verscicolor

Sepal Length	Sepal Width	Species
5.2	3.1	?

Step 1: Find Distance

Distance (Sepal Length, Sepal Width) = $\sqrt{(x-a)^2 + (y-b)^2}$

Distance (Sepal Length, Sepal Width) = $\sqrt{(5.2-5.3)^2 + (3.1-3.7)^2}$

Distance (Sepal Length, Sepal Width) = 0.608

Sepal Length	Sepal Width	Species	Distance
5.3	3.7	Setosa	0.608

KNN Classifier Solved Example - 1

Sepal Length	Sepal Width	Species	Distance	Rank
5.3	3.7	Setosa	0.608	3
5.1	3.8	Setosa	0.707	6
7.2	3.0	Virginica	2.002	13
5.4	3.4	Setosa	0.36	2
5.1	3.3	Setosa	0.22	1
5.4	3.9	Setosa	0.82	8
7.4	2.8	Virginica	2.22	15
6.1	2.8	Verscicolor	0.94	10
7.3	2.9	Virginica	2.1	14
6.0	2.7	Verscicolor	0.89	9
5.8	2.8	Virginica	0.67	5
6.3	2.3	Verscicolor	1.36	12
5.1	2.5	Verscicolor	0.60	4
6.3	2.5	Verscicolor	1.25	11
5.5	2.4	Verscicolor	0.75	7

Step 3: Find the Nearest Neighbor

If k = 1 – Setosa

If k = 2 – Setosa

K-Means Clustering – Solved Example

Current Centroids:
A1: (2, 10)
B1: (6, 6)
C1: (1.5, 3.5)

Data Points			Distance to						Cluster	New Cluster
			2	10	6	6	1.5	1.5		
A1	2	10	0.00		5.66		6.52		1	
A2	2	5	5.00		4.12		1.58		3	
A3	8	4	8.49		2.83		6.52		2	
B1	5	8	3.61		2.24		5.70		2	
B2	7	5	7.07		1.41		5.70		2	
B3	6	4	7.21		2.00		4.53		2	
C1	1	2	8.06		6.40		1.58		3	
C2	4	9	2.24		3.61		6.04		2	

$$d(p_1, p_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

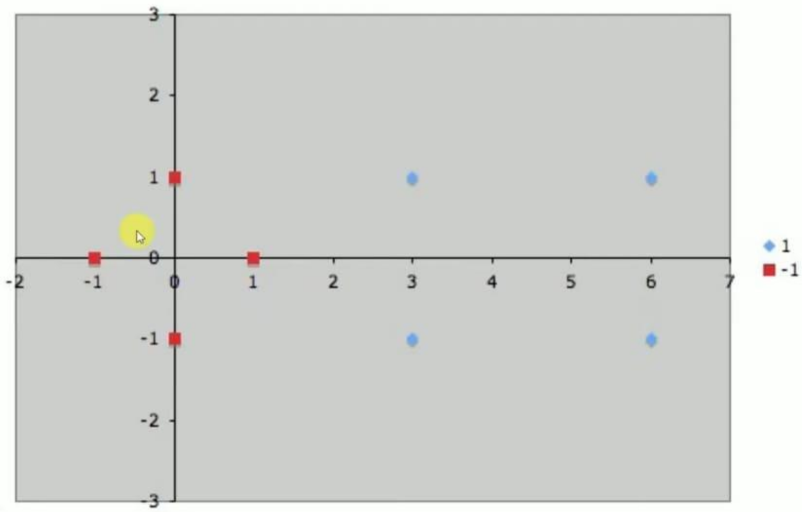
Support Vector Machine - Linear Example Solved

Suppose we are given the following positively labeled data points,

$$\left\{ \begin{pmatrix} 3 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ -1 \end{pmatrix}, \begin{pmatrix} 6 \\ 1 \end{pmatrix}, \begin{pmatrix} 6 \\ -1 \end{pmatrix} \right\}$$

and the following negatively labeled data points,

$$\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \end{pmatrix} \right\}$$



Support Vector Machine - Linear Example Solved

- Each vector is augmented with a 1 as a bias input
- So, $s_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, then $\tilde{s_1} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$
- Similarly,
- $s_2 = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$, then $\tilde{s_2} = \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix}$ *and* $s_3 = \begin{pmatrix} 3 \\ -1 \end{pmatrix}$, then $\tilde{s_3} = \begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix}$

Support Vector Machine - Linear Example Solved

$$\begin{aligned}
\alpha_1 \tilde{s_1} \cdot \tilde{s_1} + \alpha_2 \tilde{s_2} \cdot \tilde{s_1} + \alpha_3 \tilde{s_3} \cdot \tilde{s_1} &= -1 \\
\alpha_1 \tilde{s_1} \cdot \tilde{s_2} + \alpha_2 \tilde{s_2} \cdot \tilde{s_2} + \alpha_3 \tilde{s_3} \cdot \tilde{s_2} &= +1 \\
\alpha_1 \tilde{s_1} \cdot \tilde{s_3} + \alpha_2 \tilde{s_2} \cdot \tilde{s_3} + \alpha_3 \tilde{s_3} \cdot \tilde{s_3} &= +1
\end{aligned}$$

$$\alpha_1(1+0+1)+\alpha_2(3+0+1)+\alpha_3(3+0+1) = -1$$
$$\alpha_1(3+0+1)+\alpha_2(9+1+1)+\alpha_3(9-1+1) = 1$$
$$\alpha_1(3+0+1)+\alpha_2(9-1+1)+\alpha_3(9+1+1) = 1$$

$$\alpha_1 \begin{pmatrix}1\\0\\1\end{pmatrix}\begin{pmatrix}1\\0\\1\end{pmatrix} + \alpha_2 \begin{pmatrix}3\\1\\1\end{pmatrix}\begin{pmatrix}1\\0\\1\end{pmatrix} + \alpha_3 \begin{pmatrix}3\\-1\\1\end{pmatrix}\begin{pmatrix}1\\0\\1\end{pmatrix} = -1$$

$$\alpha_1 \begin{pmatrix}1\\0\\1\end{pmatrix}\begin{pmatrix}3\\1\\1\end{pmatrix} + \alpha_2 \begin{pmatrix}3\\1\\1\end{pmatrix}\begin{pmatrix}3\\1\\1\end{pmatrix} + \alpha_3 \begin{pmatrix}3\\-1\\1\end{pmatrix}\begin{pmatrix}3\\1\\1\end{pmatrix} = 1$$

$$\alpha_1 \begin{pmatrix}1\\0\\1\end{pmatrix}\begin{pmatrix}3\\-1\\1\end{pmatrix} + \alpha_2 \begin{pmatrix}3\\1\\1\end{pmatrix}\begin{pmatrix}3\\-1\\1\end{pmatrix} + \alpha_3 \begin{pmatrix}3\\-1\\1\end{pmatrix}\begin{pmatrix}3\\-1\\1\end{pmatrix} = 1$$

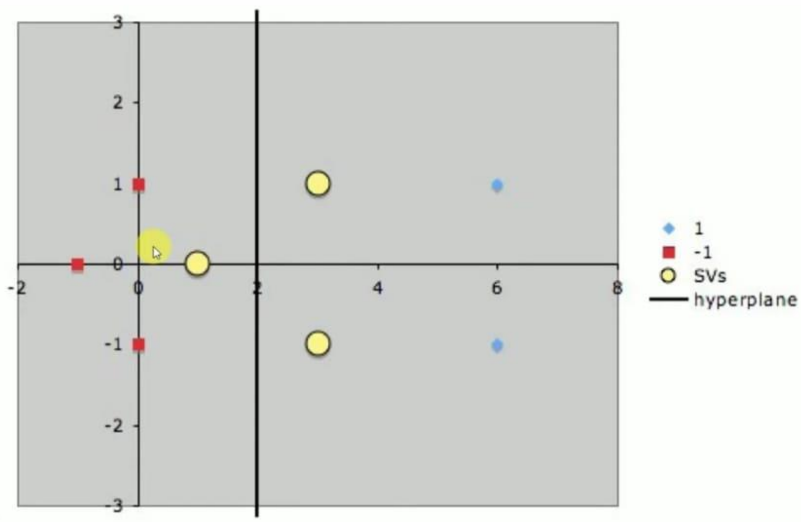
$$2\alpha_1 + 4\alpha_2 + 4\alpha_3 = -1$$
$$4\alpha_1 + 11\alpha_2 + 9\alpha_3 = 1$$
$$4\alpha_1 + 9\alpha_2 + 11\alpha_3 = 1$$

$$\alpha_1 = -3.5$$
$$\alpha_2 = 0.75$$
$$\alpha_3 = 0.75$$

Support Vector Machine - Linear Example Solved

$$\begin{aligned}\tilde{w} &= \sum_i \alpha_i \tilde{s}_i \\ &= -3.5\begin{pmatrix}1\\0\\1\end{pmatrix} + 0.75\begin{pmatrix}3\\1\\1\end{pmatrix} + 0.75\begin{pmatrix}3\\-1\\1\end{pmatrix} \\ &= \begin{pmatrix}1\\0\\-2\end{pmatrix}\end{aligned}$$

- Finally, remembering that our vectors are augmented with a bias.
- We can equate the last entry in $\widetilde{w}$ as the hyperplane offset b and write the separating
- Hyperplane equation $\boldsymbol{y} = \boldsymbol{wx} + \boldsymbol{b}$
- with $\boldsymbol{w} = \binom{1}{0}$ and $\boldsymbol{b} = -2$.



Mid-Term

Q2:

X1	X2	Y
0	0	0
0	1	0
1	0	1
1	1	1

Build decision tree Y [0=+ , 1=-]

S [+2, 2-] Entropy(S) = 1

Attribute X1:

Values: 0, 1

$S_0 [+2, 0-]$ Entropy$(S_0) = 0$

$S_1 [+0, 2-]$ Entropy$(S_1) = 0$

$$Gain(S, X1) = Entropy(S) - \sum \frac{|S_v|}{|S|} Entropy(S_v)$$

$$= 1 - \left[\frac{2}{4}(0) + \frac{2}{4}(0)\right]$$

$$Gain(S, X1) = 1 - 0 = 1$$

Attribute : X2

$S_0 [+1, 1-]$ Entropy $(S_0) = 1$

$S_1 (-1, 1+]$ Entropy $(S_1) = 1$

$$Gain(S, X2) = Entropy(S) - \sum \frac{|S_v|}{|S|} Entropy(S_v)$$

$$= 1 - \left[\frac{2}{4}(1) + \frac{2}{4}(1)\right]$$

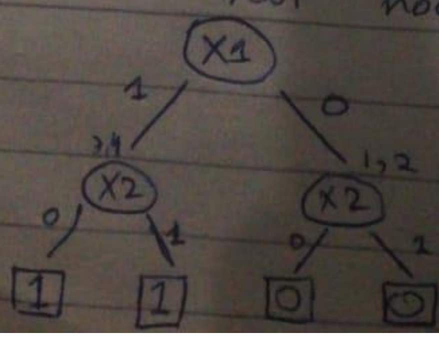
$$Gain(S, X2) = 1 - 1 = 0$$

Comparing Gain :

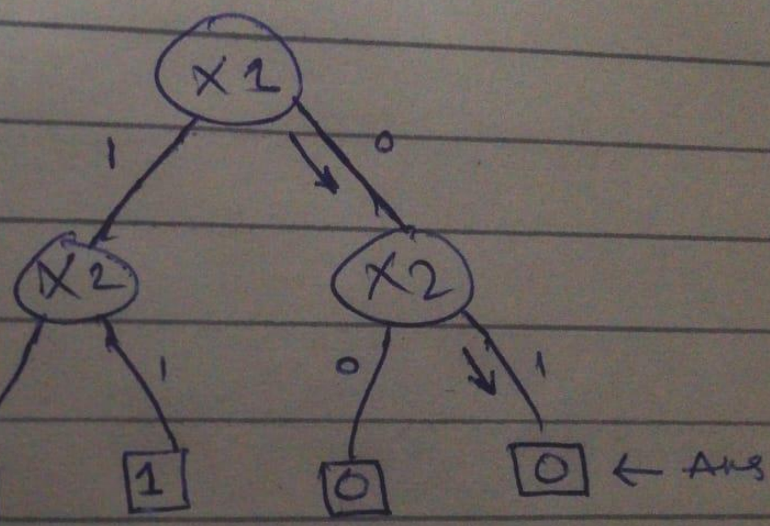
$Gain(S, X1) = 1$

$Gain(S, X2) = 0$

X1 becomes root node.



Classify $X1 = 0$, $X2 = 1$ on Tree:-



When $X1 = 0$, $X2 = 1$ then $\boxed{Y = 0}$