NPTEL Week-6 Live Session

on Machine Learning and Deep Learning - Fundamentals and Applications (noc24_ee146)

A course offered by: Prof. Manas Kamal Bhuyan, IIT Guwahati

NPTEL Quiz Solution: Week-5; SVM implementation

Powered by:

PMRF

Prime Minister's Research Fellows
Ministry of Education
Government of India

By

Arka Roy
NPTEL PMRF TA

Prime Minister's Research Fellow
Department of Electrical Engineering, IIT Patna
Web: https://sites.google.com/view/arka-roy/home

1) Which of the following statements is true about logistic regression?

- It is robust to extreme outliers in the data
- It is not suitable for binary classification problems
- It does not make any assumptions about the data distribution
- It requires a large amount of training data to perform well

Linear Regression:- To fit a curve.

Prediction $\Rightarrow \hat{y} = w^T x + b$.

$\underbrace{w \text{ (Weights)} \quad b \text{ (Bias)}}_{\text{Learnable parameters.}}$

Dataset $\{x_i, y_i\}$ — y_i: True output

$\mathcal{L}(y_i, \hat{y}_i) = MSE(y_i, \hat{y}_i)$

$\mathcal{L} = \frac{1}{2} \sum_{i=1} (y_i - \hat{y}_i)^2$

Input $(x) \in \mathbb{R}$. (Real numbers)

Output $(\hat{y}) \in \mathbb{R}$.

Logistic regression:-
$\Downarrow$
classification

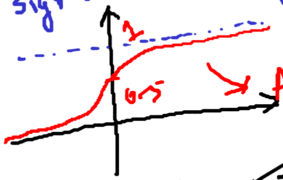
$\{x_i, y_i\}$ → only discrete class labels.

x_i: Feature

y_i: Label → Type → Discrete.
→ Categorical Variable (string data)

Example:- → "Non-Diabetes"
→ "Diabetes"
→ Binary classification → Discrete / Categorical dependent Variable

sigmoid



$y_i \in \mathbb{Z}$

→ Assymptotic curve

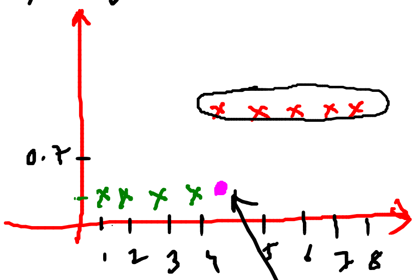
$$\hat{y} = \frac{1}{1 + e^{-(w^T x + b)}}$$

Sigmoid function

$\hat{y} \in (0, 1)$ → Binary classification

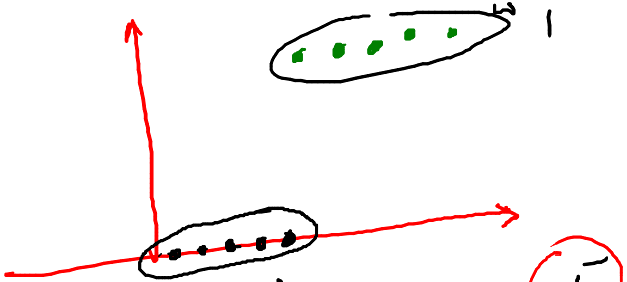
1) Which of the following statements is true about logistic regression?

- ~~It is robust to extreme outliers in the data~~ False.
- ~~It is not suitable for binary classification problems~~ False.
- ✓ It does not make any assumptions about the data distribution
- ~~It requires a large amount of training data to perform well~~ False.



Cat - Dog.

extreme outlier



$$y_i = \frac{1}{1+e^{-(w^T x_i + b)}}$$

$\geq 0.5 \rightarrow w_1$

$< 0.5 \rightarrow w_2$

→ Binary classification

→ Even with smaller amount of data you can train a logistic regression

→ Data output range $\in (0,1)$

→ Data output ⇒ Discrete/categorical dependent variable.

→ Do not concern Pre notion about specific prob. dist. function.

LDF, NB -

$x \in N(0,1) \rightarrow$ univariate feature.

x_2

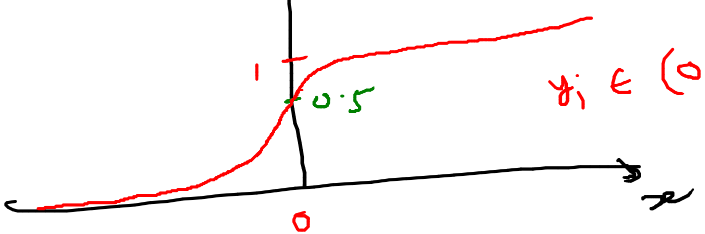
2) Logistic regression is a machine learning algorithm that is used to predict the probability of a _?

- Categorical independent variable
- Categorical dependent variable
- Numerical independent variable
- Numerical dependent variable

3) What's the the hypothesis of logistic regression?

- ◯ To limit the cost function between 0 and 1
- ◯ To limit the cost function between -1 and 1
- ◯ To limit the cost function between -infinity and +infinity
- ◯ To limit the cost function between 0 and +infinity

$\hat{y}(x) =$ $\quad y = \frac{1}{1+e^{-x}}$



$y_i \in (0,1)$

4) The dataset of pass or fail in an exam for five students are given in the table.

Hours Study	Pass(1) / Fail (0)
29	0
15	0
33	1
28	1
39	1

Calculate the probability of pass for student who studies 34 hours.

$P(\text{Pass})| = ?$ for hrs = 34

Assume the model suggested by the optimizer for odds of passing the course is:

→ log(odds) = - 64 + 2 x hours

- 0.932
- 0.952
- ✓ 0.982
- 0.992

$P(\text{Pass}) = 0.9820$

$$\log(\text{odd}(\text{Pass})) = -64 + (2 \times \text{hrs})$$
$$= -64 + (2 \times 34)$$
$$\log(\text{odd}(\text{Pass})) = -64 + 68 = 4$$
$$\boxed{\text{odd}(\text{Pass}) = e^4} \quad (1)$$

Event = A. Trials = N Occurrence of event A $= N_A$

$$P(A) = \frac{N_A}{N}$$

$$\text{odd}(A) = \frac{\text{Probability of event A}}{\text{Probability of non occurrence of event A}}$$

$$\boxed{\text{odd}(A) = \frac{P(A)}{1 - P(A)}} \quad (A)$$

$$\text{odd}(\text{Pass})[1 - P(\text{Pass})] = P(\text{Pass})$$
$$\Rightarrow P(\text{Pass})(1 + \text{odd}(\cdot)) = \text{odd}(\cdot)$$
$$\Rightarrow P(\text{Pass}) = \frac{\text{odd}(\text{Pass})}{1 + \text{odd}(\text{Pass})}$$
$$P(\text{Pass}) = \frac{e^4}{1 + e^4} = \frac{1}{1 + e^{-4}}$$

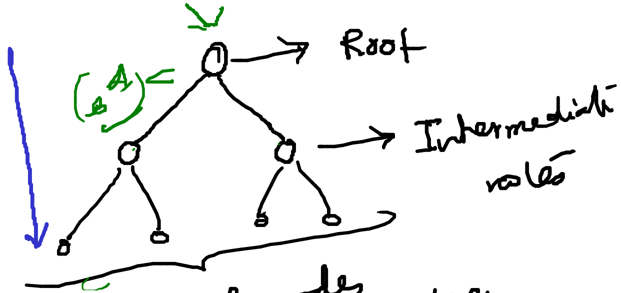
5) Which of the following is used to measure the quality of a split in a decision tree?

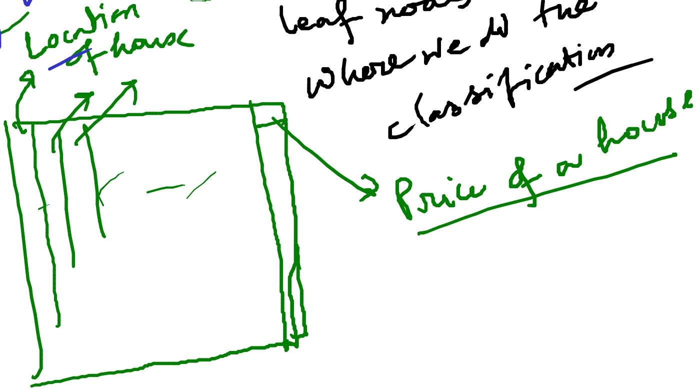
- Information gain
- Gini index
- Entropy
- ~~All~~ of the above

Measure | Quality measures

By observing them (↑) you do the branching of decision Tree

Entropy,
Gini Index.
Information gain





6) There are six instances for attributes a1 and a2, with positive and negative classification results given below:

Instance	Classification	a1	a2
1	+	T	T
2	+	T	T
3	-	T	F
4	+	F	F
5	-	F	T
6	-	F	T

What is the information gain of a1 relative to these training instances?

- 0.0523
- ✓ 0.0817
- 0.025
- 0

$$\boxed{\text{Information gain} = \text{Entropy}(S) - \sum \frac{|S_i|}{|S|} \cdot \text{Entropy}(S_i)}$$

We are focusing on input attribute A_1

$A_1 \in T/F \cong S_1, S_2$

$$\text{Entropy}(T)\Big|_{A_1} = -\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3} = 0.9183$$

$$\text{Entropy}(F)\Big|_{A_1} = -\frac{1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log\frac{2}{3} = 0.9183.$$

$\sum -p(x)\log_2 p(x)$

$$\text{Entropy(whole set)} = -\frac{3}{6}\log_2\frac{3}{6} - \frac{3}{6}\log_2\frac{3}{6} = 1.$$

$$\text{IG} = \text{Entropy(whole set)} - \left(\frac{|S_T|}{|S_{whole}|} \cdot \text{Entropy}(T)\Big|_{A_1} + \frac{|S_F|}{|S_{whole}|} \cdot \text{Entropy}(F)\Big|_{A_1}\right)$$

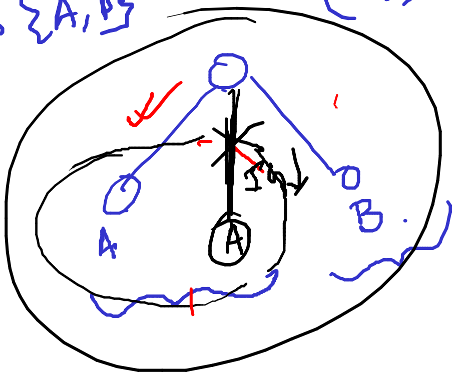
$$= 1 - \left(\frac{3}{6} \times 0.9183 + \frac{3}{6} \times 0.9183\right) = 0.0817. \text{ (Ans)}$$

7) How is pruning used to prevent overfitting in decision trees?

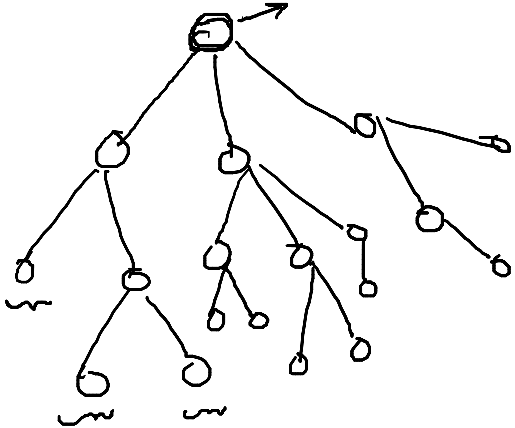
- By limiting the depth of the tree
- By removing branches with low information gain
- By setting a minimum number of instances per leaf
- By limiting the maximum number of leaves

Three (x_i) (x_1, x_2, x_3)

classes {A, B}

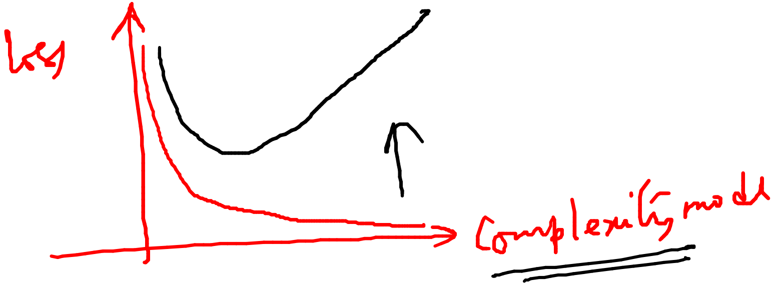


After training of DT you have got this structure



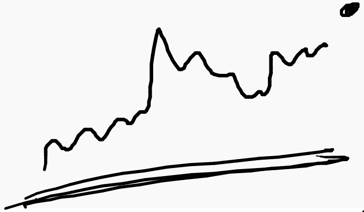
reduced complexity

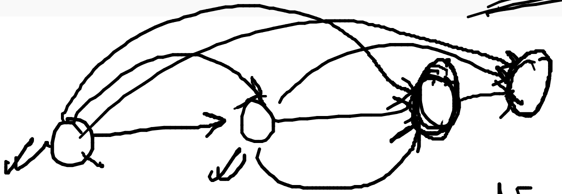
Reduced Redundant information



8) Where does the additional variables are added in HMM?

- Temporal model
- Reality model
- Probabilistic model
- All the above





x_t dependent on x_{t-k}

Temporal data notion.

find optimal Decision Boundary.

Linear

Nonlinear SVM

Polynomial SVM

Cubic SVM

Quadratic SVM

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0

XOR gate