AAI - 695 Mid-Term Exam

Chinmay Bhagwat CWID → 20015512

1) Bias - variance trade-off.

Bias - variance trade off is one of the common problem in supervised machine learning. Bias is difference between the average prediction of the model & the actual value. model with high bias lead to underfitting and gives testing error. Variance is the variability of spread of model prediction.

Trade off is tension between the error introduced by bias & variance. This trade off in complexity results trade off in bias & variance. Bias & variance are inversely ~~coreted~~ corelated. So that both can't be high or low at the same time.

- To reduce the variance we can increase training set data & Increasing the lambda value also work. If we reduce the number of feature in the model it will reduce variance.
- By increasing the features, decreasing the alpha parameters & by performing feature engineering we can reduce bias.

1.2) Over fitting.

Overfitting is the issues ~~occur~~ in machine learning model which occurs when the model gives accurate prediction of for training data but not for new data. when the model learns training data too well, overfitting will occurs.

How to reduce overfitting.

- Cross validation
- Regularization
- Remove feature
- Early stopping & ensembling.

1.3) Learning rate.

Learning rate is the parameter which defines how quickly we are moving towards optimal weights..

Learning rate cannot be too large or too small because if learning rate is set too high it will cause divergent behaviour in loss function. & if its too low training will goes too slow.

1.4) Multi class classifier from binary classifier.

we can use 3 methods to make ~~multy~~ multi-class classifier from binary classifier

→ One-Vs-Rest :- In here it will use multiple rounds of binary classification for multiclass classification.

→ One-Vs-One :- It construct a binary classifier for each pair of classes.

→ Error-correcting Output Codes :- Here it encodes K classes into N bit vector. each classes are represented as a bit of each vector.

+)

.5) We initialize each weight to very small number because it is an ~~exp~~ expectation of ~~stoch~~ stochastic optimization algorithm used to train model to receive very small value.

2.1)

So here $TP = 70$

$FN = 30$

$FP = 50$

$TN = 70.$

Precision :- $\frac{TP}{TP+FP} = \frac{70}{120} = \underline{0.58}$

Recall :- $\frac{TP}{TP+FN} = \frac{70}{70+30} = \underline{0.7}$

F-1 score $\rightarrow$ $\frac{2\,(\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} = \underline{0.63}$

3.1) Entropy (S) :-

$-\frac{2}{4}\log_2\left(\frac{2}{4}\right) - \frac{2}{4}\log_2\frac{2}{4} = 1$ $\left\{ \begin{array}{l} P(Y) \rightarrow 2/4 \\ P(N) = 2/4 \end{array} \right.$

3.2) Attribute sky over S.

$E(\text{cloudy}) = -\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right) = 1$

$E(\text{sunny}) = -\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right) = 1$

$\text{Gain}(S, \text{sky}) = 1 - 2/4 \times 1 - 2/4 \times 1$

$= \underline{0}$

~~Gain Entropy~~ For Temp

$E(\text{low temp}) = -1/1 \log_2(1/1) - 0 = 0.$

$E(\text{high}) = -1/3 \log_2(1/3) - 2/3 \log_2(2/3) = 0.918$

$\text{Gain}(S, \text{temp}) = 1 - 0 - 3/4(0.91) = \underline{\underline{0.3175}}$

For wind

$E(\text{mild}) = -1/2 \log_2(1/2) - 1/2 \log_2(1/2) = 1$

$E(\text{strong}) = -1/2 \log_2(1/2) - 1/2 \log_2(1/2) = 1$

$\text{Gain}(S, \text{wind}) = 1 - 2/4(1) - 2/4(1) = \underline{\underline{0.}}$

3.3) From the above calculation we can see that temperature gain is ie we will get option "C" as a result.

3.4) The last two nodes of the tree can be interchangeable. Here tree having 100% accuracy. But we can't go further without changing accuracy.

4) Outlook

	Y	N
Sunny	1/6	3/4
Overcast	2/6	0
Rain	3/6	1/4

Wind

	Y	N
Weak	5/6	2/4
Strong	1/6	2/4

$P(yes) = 6/10$

$P(NO) = 4/10$

Temperature

	Y	N
Hot	1/6	2/4
mild	2/6	1/4
cool	3/6	1/4

Humidity

	Y	N
High	2/6	3/4
Normal	4/6	1/4

For largest value (Yes) = $P(yes) \times P(Rain/yes) \times P(Hot/yes) \times P(High/yes) \times P(weak/yes)$

$= 6/10 \times 3/6 \times 1/6 \times 2/6 \times 5/6$

$= 0.0138$

For largest value (NO) = $P(NO) \times P(Rain/NO) \times P(Heat/NO) \times P(High/NO) \times P(weak/NO)$

$= 4/10 \times 1/4 \times 2/4 \times 3/4 \times 2/4$

$= 0.018$

Therefore instance will be "NO"

5)

5.1) A → AND

B → OR

C → NAND.

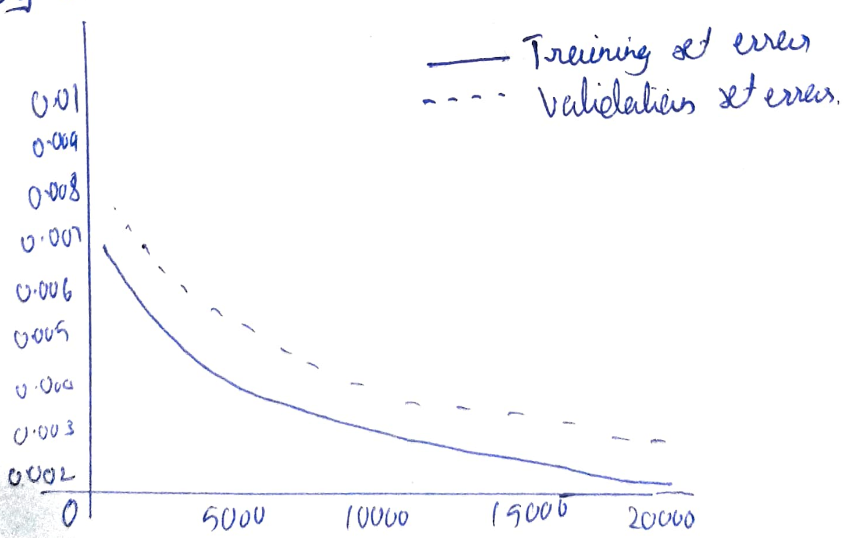
5.2) All the primitive boolean functions AND, OR, NAND, NOR can be represented ~~as~~ by perceptrons.

Any boolean function can be implemented by using a (2 level) combination of these primitives. this is called functional completeness property.

6.1) Yes, Overfitting can be observed in this graph.

we can detect overfitting by ~~used~~ visualizing the training set error & validation set error. If the gap between these two lines are high then we can consider it as overfitting.

6.2) when ~~we~~ we double the size of training data, overfitting will reduce. when we double the size model learn new information about the data. That's why it able to generalize the behaviour of the testing data.



7.1) class 2 have 2 votes & class 1 have 1 votes
Therefore <u>class 2</u> is the final decision.

7.2) For class 1 → classifies 3 → $0.2 \times 1 = 0.2$

For class 2 → classifies 1 & 2 → 0.8.

Therefore <u>class 2</u> is the final decision.

7.3) For class 1

$\hat{p}(w_1/d_{1,2}(x)=1) = \frac{30}{30+50} = 0.375$

$\hat{p}(w_1/d_{2,2}(x)=1) = \frac{40}{40+60} = 0.4$

$\hat{p}(w_1/d_{3,1}(x)=1) = \frac{80}{80+70} = 0.533$

For class 2

$\hat{p}(w_2/d_{1,2}(x)=1) = \frac{50}{30+50} = 0.62$

$\hat{p}(w_2/d_{2,2}(x)=1) = \frac{60}{40+60} = 0.6$

$\hat{p}(w_2/d_{3,1}(x)=1) = \frac{70}{80+70} = 0.46$

w(class 1) = 0.0799

w(class 2) = 0.171

Hence <u>class 2</u> is selected.