Mid-I Review.

2. point estimator $\quad MSE = (Bias)^2 + Var$

$Bias = \theta - E(\hat{\theta}) \qquad Var(\hat{\theta}) = E(\theta - E(\hat{\theta}))^2)$

$E[X] = \frac{1}{n}\sum_{i}^{n} X_i$

3. X_train, X_test, y_train, y_test = train_test_split(x, y, testsize =0.3)

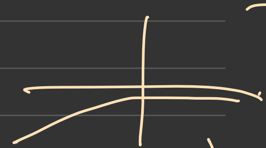
```
model = models.Sequential()
 model.add(...)
 :
 :
  model.compile(optimizer = optimizers.RMSprop(lr=0.001),
  loss = 'binary_crossentropy', metrics= ['accuracy'])

history = model.fit( X_train, y_train, epochs=20, batch_size=512,
   validation_data= (X_test, y_test)) → problem.
  y_pred = model.predict (X_test)
```

We are using test set for validation set.

we need to split the train data into 80 : 20 as x_sub train and x_val data.

4. Assume

$L(\theta, y) = -\log\left(\frac{1}{1+e^{(1-2y)\theta}}\right)$

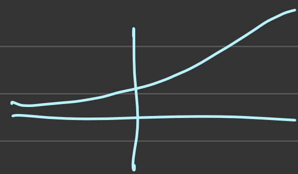


$\log\left(\frac{1}{1+e^{\theta}}\right)$

for $y = 0 \Rightarrow -\log\left(\frac{1}{1+e^{\theta}}\right)$ → soft plus function

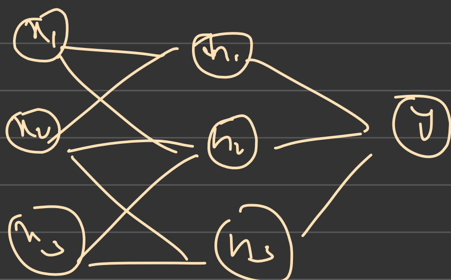
→ fipping y axis as their is -

The curve



To minimize → Decreasing

(5)



$$\frac{\partial L}{\partial x_1} \qquad y = h_1 + h_2 + h_3$$

$h_1 = x_1 x_2 \quad h_2 = x_1 x_3 \quad h_3 = x_2 x_3$

$$\frac{\partial L}{\partial x_1} = \frac{\partial L}{\partial y}\left(\frac{\partial y}{\partial h_1}\frac{\partial h_1}{\partial x_1} + \frac{\partial y}{\partial h_2} \cdot \frac{\partial h_2}{\partial x_1}\right)$$

$$= \frac{\partial L}{\partial y}(x_2 + x_3)$$

(6).

(c) → $L = -\sum_k y_k \log(o_k)$ $\quad y \in \{0, 1\}$ is one hot encoded class label.

value of $\frac{\partial L}{\partial x_i}$?

$$\frac{\partial L}{\partial z_i} = \sum_j \frac{\partial L}{\partial o_j}\frac{\partial o_j}{\partial z_i} \quad \text{and} \quad \sum_k y_k = 1$$

$\Rightarrow y \begin{cases} 0 \\ 1 \end{cases}$ then O must be $y=0$ then $O_i=0$

$y=1$ then $O_i=1$

Substitute the values in fn & check

$O_i - y_i$

if $O_i = 0.5$ and $y=0$ then $O_i - y_i = 0.5$

then here we need to decrease O_i to $0 \downarrow$

$\Rightarrow$ $y_i = 0$ $O_i = 0$ to decrease Loss function

if $O_i = 0.5$ and $y = 1$ then $O_i - y_i = -0.5$

then we need to increase O_i to -1 to make

$\Rightarrow$ $y_i = 1$ $O_i =$ Loss function

0

try with different O_i values.

a) $\theta_i = \dfrac{e^{zk}}{\sum_k e^{zk}} = e^{zk} \cdot \left(\sum_k e^{zk} \right)^{-1}$

Find $\dfrac{\partial o_i}{\partial z_j}$ when $i \neq j$

Ans. $o_i = \dfrac{e^{z_i}}{e^{z_0} + e^{z_1} \cdots + e^{z_i} + e^{z_j} + \cdots}$

when $i \neq j$ we are finding ∂z_j then e^{z_i} is constant.

$= e^{z_i} - 1 \left(\sum e^{z_k} \right)^{2} \cdot e^{z_j}$

$\dfrac{d}{dx} \left(\dfrac{1}{e^u + 1} \right) = -1 \dfrac{1}{(e^u + 1)^2} \cdot e^u$

$\Rightarrow e^{z_i} - 1 \left(\sum e^{z_k} \right)^{-2} \cdot e^{z_j} = \dfrac{e^{z_i}}{\sum e^{z_k}} \cdot \dfrac{e^{z_j}}{\sum e^{z_k}} = o_i \, o_j$

When $i = j$ $\quad f(u) \cdot g(u) = f' g(u) + g' f(u)$

$\Rightarrow e^{z_i} \dfrac{1}{\sum e^{z_k}} + e^{z_i} \dfrac{-1}{(\sum e^{z_k})^2} \cdot e^{z_i}$

$$= \frac{e^{z_i}}{\sum e^{z_k}} \left(1 - \frac{e^{z_i}}{\sum e^{z_k}}\right) \quad = \quad (o_i)(1 - o_j)$$

$$(o_i)\ (1 - o_i)$$

o_i $\quad$ o_j or o_i

6.3

3 Ans

$$\rightarrow \frac{\partial L}{\partial z_i} = \sum_j \frac{\partial L}{\partial o_j} \cdot \frac{\partial o_j}{\partial z_i}$$

$$\sum_{j \neq i} \frac{\partial L}{\partial o_j} \cdot \underbrace{\frac{\partial o_j}{\partial z_i}}_{-o_i o_j} + \frac{\partial L}{\partial o_i} \cdot \underbrace{\frac{\partial o_i}{\partial z_i}}_{o_i(1-o_i)}$$

$$\left(\frac{\partial L}{\partial o_j} = -y_j \frac{1}{o_j}\right)$$

$$= \sum_{j \neq i} -y_j \frac{-o_i o_j}{o_j} + -y_i \frac{1}{o_i} o_i (1 - o_i)$$

$$= \sum_{j \neq i} y_j o_i \quad \underbrace{- y_i (1 - o_i)}_{-y_i + y_i o_i}$$

$$= \sum_{j \neq i} y_j o_i - y_i \quad = \quad o_i \sum_j y_j - y_i$$

$$\rightarrow \sum_j y_j = 1 \text{ then } \quad o_i - y_i$$

T or F

→ a. False → classification & Regressions

b. True - Supervised learning

c. False - data it has not seen before

d. False : overfitting - underfitting

e. True : $y = \sum_{k=1}^{d} x_i$

? f)

? g) True - XOR cannot be solved using logistic

! $\overline{h)}$? i)

j) True - one hidden layer is enough

k) True - Bayes error.