Formulas

UNIT I

1. Marginal Probability

$$P(X=x) = P(X,Y) + P(Y,\bar{Y})$$

2. Conditional Probability

$$P(X/Y) = \frac{P(X \wedge Y)}{P(Y)}$$

3. Chain rule

$$P(A_1 \cdots A_n) = P(A_1)\, P(A_2/A_1)\, P(A_3/A_1 \cdot A_2) \cdots P\left(\frac{A_n}{A_1 \cdots A_{n-1}}\right)$$

4. Independent

$$P(A \cap B) = P(A) \cdot P(B)$$

5. Conditional Independence

$$P(A/B, C) = P(A/C)$$

i.e A is independent of occurance of B

6. Expectation $E(x) = \Sigma x P(X)$

7. Variance $V(x) = \sigma^2(x)$

$$\sigma^2(x) = E(x^2) - [E[x]]^2$$ $\sigma(x)$ is standard deviation

8. Covariance

$$Cov(x,y) = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})$$

9. Correlation

$$\mathrm{Corr}(x,y) = \frac{\mathrm{cov}(x,y)}{\sigma x \sigma y}$$

10. Linear Regression

$$m = \frac{N \times \Sigma(xy) - \Sigma x \Sigma y}{N\,\Sigma(x^2) - (\Sigma(x))^2}$$

$$b = \frac{\Sigma y - m\Sigma x}{N}$$

$$y = mx + b$$

11. $$\mathrm{cost} = \frac{1}{2m} \Sigma (y^i - \hat{y}(i))^2$$ m = no of observations

w_0

$$w_{new} = w_0 - \alpha\left(\frac{\partial J}{\partial w_0}\right)$$

$$= w_0 - \alpha \Sigma\left[\frac{-1}{m}(y^i - \hat{y}^{(i)})\right]$$

w_1

$$w_{new} = w_0 - \alpha \Sigma\left[\frac{-1}{m}(y^i - \hat{y}(i))(x_i)\right]$$

$$y = w_0 + w_1 x$$

UNIT 2

1. Bayes theorem

$$P(C_i/x) = \frac{P(x/C_i)\,P(C_i)}{P(x)}$$

2. Sigmoid function

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

$$\frac{\partial \sigma(x)}{\partial(x)} = \sigma(x)(1-\sigma(x))$$

3. SVM

$X \rightarrow$ data points

$Y \rightarrow$ target vector t_n

Steps

1. Find XY, $(XY)^2 = (XY)(XY)^T$
2. Dual optimization problem from matrix in terms of $(a_1 \ldots a_n)$ as

$= (a_1 + a_2 + \ldots + a_n) - \frac{1}{2}$ (Values from matrix) — (2)

3. $\sum_{n=1}^{N} a_n t_n = 0$ (find relation between a_is to subs in (2))

4. After (2) has reduced to certain variables perform partial derivative wrt to $a_1 \ldots a_i$

find values of $a_1 \ldots a_n$

5. Values > 0 are real support vectors

6. Find W using $W = \sum_{n=1}^{N} a_n t_n x_n$

7. Norm of w is $\|w\|$ is the margin

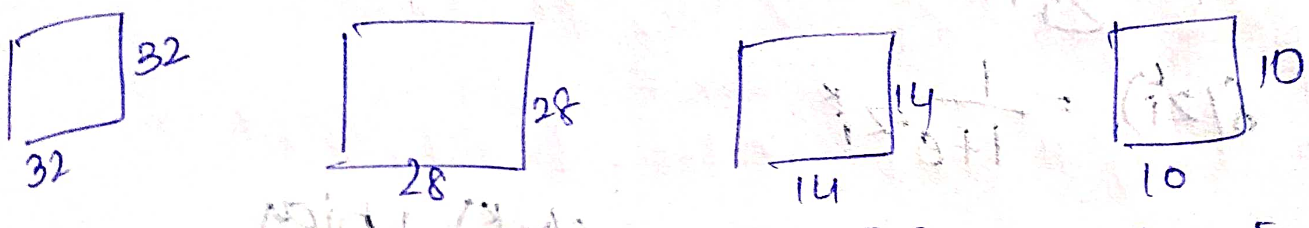
choose any input data point vector for finding threshold

$t_i(Wx_i - b) = 1$ find b

UNIT 3

1. CNN - finding no of Parameters required in each layer.

Input	Conv1	Pooling	Conv2



	S=1 F=5	S=1 F=2	S=1 F=5
	K=6 P=10	K=6 P=10	K=16 P=0
32X32X3	28X28X6	14X14X6	10X10X6

(1) Input layer : no of params = 0

= ((shape width × Height of filter * number of filters in previous layer +1) * no of filter

= ((5 * 5 * 3 + 1)) * 6

= 456

② Pool layer : no learning parameters ∴ params = 0

③ $Conv2D = ((5*5*6)+1)*16$

$= (150+1)16$

$= 2416.$

Activation shape (width, height, dim of input)

Activation size $w*h*D$

④ For fully connected layers / Softmax layer.

no of params = ((current layer neurons (will be given) * previous layer neurons (activation size) + (1 * current))

Back Propogation

1. Forward propogation

$$z_i^K = \Sigma \text{Transition} \times \text{hitting} + \text{bias}$$

$$\delta(z_i^K) = \frac{1}{1+e^{-z_i^K}}$$

$$y^K = \Sigma \text{Transition} \times \delta(z^K) + \text{bias}$$

$$\delta(y_i^K) = \frac{1}{1+e^{-y_i^K}}$$

2. Back Propogation

~~Hi~~ Error calculation

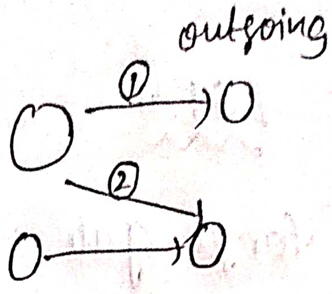
$$\varepsilon_i = \text{given} - \delta(y^K)$$

Hidden to output

1. $\delta_i = e_i^1 \, \delta(y_i^K)(1 - \delta(y_i^K))$

$W_{oldnew} = W_{old} + \eta \, \delta_i \, \delta(z_i)$

for bias $\delta(z_i) = 1$



Input to hidden

2. $\delta H_i^K = \delta(z_i^K)(1 - \delta(z_i)^K)$

$$\left[\delta_i^1 \, ①^{Trans} + \delta_i^2 \, ②^{Trans}\right]$$

$m_{new} = m_{old} + \eta \, \delta H_i^K x_i$ (x_i is input)

here also for bias $x_i = 1$

Unit 4

1. Relu function : $y = \max(0, x)$
2. Gradient ascent : $y = \eta x$
3. RNN for classification

$$h_i = \tanh(x_i U_j + h_{i-1}^{j} V_j)$$

$Y = h_n^2 W$ take softmax

$$\text{softmax} = \frac{e^x}{\sum_{i=0}^{K} e^i} \qquad e = 2.718$$

4) RNN

No of weights in Weight Matrix (input U) = Input × hidden l

Number of weights in Weight Matrix (V) = (hidden)^2

Number of weights in Weight Matrix (W - output) = hidden × no. of classes

5. LSTM

forget gate $f_t = \sigma(W_f [h_{t-1}, x_t] + b_f)$

Input gate $i_t = \sigma(W_i [h_{t-1}, x_t] + b_i)$

Output gate $O_t = \sigma(W_o [h_{t-1}, x_t] + b_o)$

New cell content $\tilde{C}_t = \tanh(W_c [h_{t-1}, x_t] + b_c)$

cell state $C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$

Hidden State $h_t = O_t * \tanh(C_t)$

6. GRU

Update gate $z_t = \sigma(W_z \cdot [h_{t-1}, x_t])$

reset gate $r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$

New hidden state content

$$\tilde{h}_t = \tanh(W [r_t * h_{t-1}, x_t])$$

Hidden state

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$