NAME: Anil Pavuluru
CSUID: 2782551
CIS-694- Assignment-4 (may. 01-2021)

1) statistical machine learning slide number-15

Entropy: $H(x) = -\sum_{i=1}^{n} P(x_i)\log P(x_i)$

maximum uncertainty/entropy is obtained by equal distribution
minimize entropy means to minimize the uncertainity.

$P = [9/25, 12/25, 4/25, 1/8, 1/10] \Rightarrow -1 * \text{Sum}(P.*\log 2(P));$

$H(x) = -[(9/25 * \log_2^{(9/25)}) + 12/25 * \log_2^{(12/25)} + 4/25 * \log_2^{(4/25)} + 1/8 * \log_2^{(1/8)} + \log_2^{(1/10)} * \frac{1}{10})]$

$= -[(9/25 * -1.473) + (12/25 * -1.0588) + (4/25 * -2.6438) + (1/8 * -3) + (\frac{1}{10} * -3.3219)]$

$= -[-0.53 - 0.50784 - 0.42304 - 0.375 - 0.3322]$

$= -[-2.1688]$

$= 2.169$

$\log_2^{9/25} = \frac{\log(9/25)}{\log(2)} = -1.473$

$\frac{\log(12/25)}{\log(2)} = -1.0588$

$\frac{\log(4/25)}{\log(2)} = -2.6438$

$\frac{\log(1/8)}{\log(2)} = -3$

$\frac{\log(1/10)}{\log(2)} = -3.3219$

Entropy of $P = [9/25, 12/25, 4/25, 1/8, 1/10]$

$\boxed{P = 2.169}$

$Q = [1/5, 1/6, 1/5, 1/8, 1/5]$ $\quad \frac{1}{5} * \log_2\left(\frac{1}{5}\right)$

Entropy of $Q = -\left[1/5 * \log_2\left(\frac{1}{5}\right) + \frac{1}{6} * \log_2\left(\frac{1}{6}\right) + \frac{1}{8} * \log_2\left(\frac{1}{8}\right) + \frac{1}{5} * \log_2\left(\frac{1}{5}\right)\right]$

$\frac{\log\left(\frac{1}{5}\right)}{\log(2)} = -2.3219$ because $\frac{1}{5}$ repeated two times.

$\frac{\log\left(\frac{1}{6}\right)}{\log(2)} = -2.5849$

$\frac{\log\left(\frac{1}{8}\right)}{\log(2)} = -3$

$$= -\left[\frac{1}{5} * (-2.3219) + \frac{1}{6} * (-2.5849) + \frac{1}{5} * (-2.3219) + \frac{1}{8} * (-3) + \frac{1}{5} * (-2.3219)\right]$$

$$= -[-0.4638 - 0.431 - 0.46438 - 0.375 - 0.46438]$$

$$= -[-2.19914]$$

$$= 2.199$$

Entropy of $Q = \left[\frac{1}{5}, \frac{1}{6}, \frac{1}{5}, \frac{1}{8}, \frac{1}{5}\right]$

$\boxed{2.199 > 2.169}$ $\quad \boxed{Q = 2.199}$

Entropy of Q > Entropy of P

uncertainty of Q > uncertainty of P

2) class1 = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
class2 = [-2, 5.5, 6.1, 5.2, 5.1, 5, 5, 6, 5, 7.3, 5.9, 5.8, 5.1]
new testing sample x = 6 → which class does belong to

get the Gaussian distribution of class1 & class2
we required μ (mean) & δ (standard deviation)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

we can infer the relative like hood based on μ & δ

$$P(C_k|x) = \frac{P(C_k)\,P(x|C_k)}{P(x)}$$

c is the class x is the condition

⇓

$$\text{Posterior} = \frac{\text{Prior} \times \text{likehood}}{\text{evidence}}$$

GC_1 : Priority is same for each class 0.5%
GC_2 :

we use maximum a posterior we use same evidence

for $GC_1 = \frac{1+2+3+4+5+6+7+8+9+10}{10}$

$\boxed{\mu_{c_1} = 5.0}$

standard deviation = $(1-5)^2 \Rightarrow 16$

$(2-5)^2 = 9$
$(3-5)^2 = 4$
$(4-5)^2 = 1$
$(5-5)^2 = 0$
$(6-5)^2 = 1$
$(7-5)^2 = 4$
$(8-5)^2 = 9$
$(9-5)^2 = 16$
$(10-5)^2 = 25$

$$\sum_{i=1}^{N} (x_i - \mu)^2$$

$$\Rightarrow \frac{16+9+4+1+0+1+4+9+16+25}{10}$$

$$\Rightarrow \frac{85}{10} = 8.5$$

$$\delta = \sqrt{\sum_{i=1}^{N} (x_i-\mu)^2} = \sqrt{8.5}$$

$\boxed{\delta_{c_1} = 2.91}$

Same for class 2 mean & standard deviation

$$GC_2 = \frac{-2+5.5+6.1+5.2+5.1+5+5+6+5+7.3+5.9+5.8+5.1}{13}$$

$$= \frac{65}{13}$$

$$= 5$$

$\sigma C_2 = (-2-5)^2 = 49$

$(5.5-5)^2 = 0.25$

$(6.1-5)^2 = 1.21$ $\qquad \sigma C_2 = \frac{1}{N}(x_i-\mu)^2$

$(5.2-5)^2 = 0.04$

$(5-5)^2 = 0$ $\qquad \Rightarrow \frac{49+0.25+1.21+0.04+0+0+1+0+5.29+0.81+0.64+0.01}{13}$

$(5-5)^2 = 0$

$(6-5)^2 = 1$

$(5-5)^2 = 0$ $\qquad \Rightarrow 4.48$

$(7.3-5)^2 = 5.29$

$(5.9-5)^2 = 0.81$ $\qquad \sigma C_2 = \sqrt{\frac{1}{N}\Sigma(x_i-\mu)^2}$

$(5.8-5)^2 = 0.64$

$(5.1-5)^2 = 0.01$ $\qquad = \sqrt{4.48} = 2.116$

$GC_2 = 5$

$\sigma C_2 = 2.116$

Hence $\cancel{f(x)}\ \frac{f_X}{y(x|y)} = f_{Y|X}(y|x) f_X(x) f_Y(y)$

it is clear posterior $= \frac{\text{prior} * \text{likehood}}{\text{evidence}}$

So $f_Y(y)$ doesnot depend on x it is difficult to estimate from law of probability we can do Gauss distribution Graph which involves integration & sumation

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$P_{C_1} = 0.13$ (in between 0.12 & 0.14) (below figures)

$P_{C_2} = 0.18$

So belong to class 2

from Gaussian graph below more likehood toward graph 2

so it belong to class 2 observe below graph implement in online calculator.

Kullback-Leibler Divergence: in mathematical statistics, the KL Divergence is a measure of how one Probability distribution is different from a second, reference Probability distribution

3) $KL(P,Q) = \Sigma P(x) \ln [P(x)/q(x)]$

$P(x) = [9/25, 12/25, 4/25, 1/8, 1/10]$

$Q(x) = [1/5, 1/5, 1/5, 1/5, 1/5]$.

$$KL = \frac{9}{25} * \ln\left(\frac{9}{25}\right) / \left(\frac{1}{5}\right) + \frac{12}{25} * \ln\left(\frac{12}{25}\right) / \frac{1}{5} + \frac{4}{25} * \ln\left(\frac{4}{25}\right) / \left(\frac{1}{5}\right)$$

$$+ \frac{1}{8} * \ln\left(\frac{1}{8}\right) / \left(\frac{1}{5}\right) + \frac{1}{10} * \ln\left(\frac{1}{10}\right) / \left(\frac{1}{5}\right)$$

$$= \frac{9}{25} * \ln\left(\frac{9 \times 5}{25}\right) + \frac{12}{25} * \ln\left(\frac{12 * 5}{25}\right) + \frac{4}{25} * \ln\left(\frac{4 * 5}{25}\right) + \frac{1}{8} * \ln\left(\frac{5}{8}\right)$$

$$+ \frac{1}{10} * \ln\left(\frac{5}{10}\right)$$

$$= \frac{9}{25} * \ln(9/5) + \frac{12}{25} * \ln\left(\frac{12}{5}\right) + \frac{4}{25} * \ln\left(\frac{4}{5}\right) + \frac{1}{8} * \ln(5/8) + \frac{1}{10} * \ln\left(\frac{1}{2}\right)$$

$$= \frac{9}{25} * 0.58 + \frac{12}{25} * 0.87 + \frac{4}{25} * (-0.22) + \frac{1}{8} * (-0.47) + \frac{1}{10} * (-0.69)$$

by solving above calculation i mean equation we get

$$\boxed{KL(P,Q) = 0.46345.}$$

$$KL(Q,P) = \Sigma q(x) \ln[q(x)/P(x)]$$

$$= \frac{1}{5} * \ln\left(\frac{1}{5}\right) * \frac{25}{9} + \frac{1}{5} * \ln\left(\frac{1}{5}\right) * \frac{25}{12} + \frac{1}{5} * \ln\left(\frac{1}{5}\right) * \frac{25}{4}$$

$$+ \frac{1}{5} * \ln\left(\frac{1}{5}\right) * 8 + \frac{1}{5} * \ln\left(\frac{1}{5}\right) * 10$$

$$= \frac{1}{5} * \ln(5/9) + \frac{1}{5} * \ln(5/12) + \frac{1}{5} * \ln(5/4) + \frac{1}{5} * \ln(8/5) + \frac{1}{5} * \ln(2)$$

$$= 0.2(-0.58 - 0.87 + 0.22 + 0.47 + 0.69)$$

$$\Rightarrow \boxed{-0.014 = KL(Q,P)}$$

4) one hot encoding works very well from most Problems until you get into situations where you have tens of thousands or even millions of classes.

In that case, your vector becomes really, really large and has mostly zeros everywhere and that becomes very inefficient

$$\begin{array}{c} a \rightarrow \\ b \rightarrow \\ c \rightarrow \\ d \rightarrow \\ \vdots \\ \vdots \\ \vdots \end{array}
\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}
\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}
\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}
\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}
\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

You will see later how we are dealing with these Problems using embeddings. it is nothing but simply comparing two vectors one that comes out classifiers and contains the Probabilities of your classes and the one-hot encoded vector that corresponds to your labels

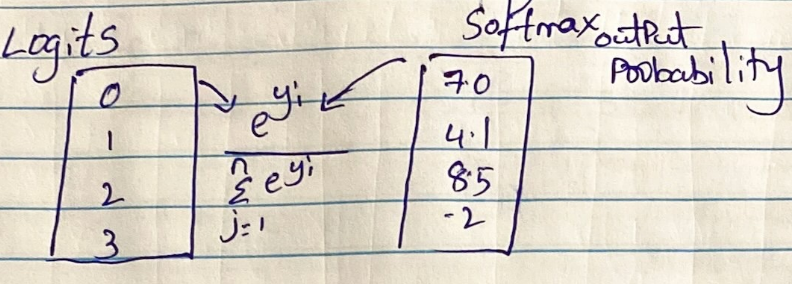
$$A \rightarrow S(y) \begin{bmatrix} 7.0 \\ 4.1 \\ 8.5 \\ -2 \end{bmatrix} \overset{?}{\longleftrightarrow} \begin{bmatrix} 1.0 \\ 0.0 \\ 0.0 \end{bmatrix}$$

<u>CROSS-entropy</u>: The natural way to measure the distance between those two Probability vectors is called the cross-entropy.

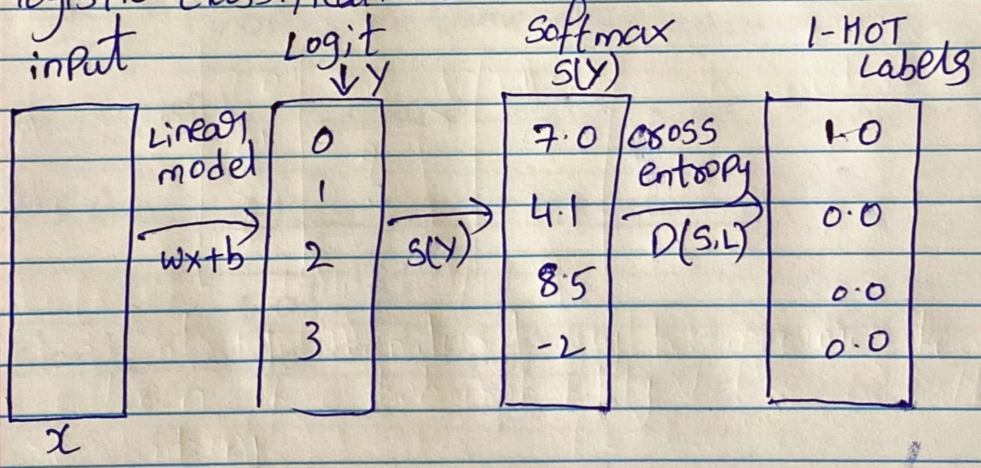
$$\underset{S(Y)}{\begin{bmatrix} 7.0 \\ 4.1 \\ 8.5 \\ -2 \end{bmatrix}} \quad \underset{\text{distance}}{D(S,L)} = -\sum_i L_i \log(S_i) \quad \underset{L}{\begin{bmatrix} 1.0 \\ 0.0 \\ 0.0 \end{bmatrix}}$$

$$D(S,L) \neq D(L,S)$$

Be careful, the cross entropy is not symmetric and you have log in that your labels and distributions should be in right Place. Your labels have a lot of zero are one-hot encoded.



So we have an input, it going to beturned into logits using a linear model, (matrix multiply & bias), we're then going to feed the logits, $wx+b$ which are scores, into softmax turn them into Probabilities and then we're going to compare those Probabilities to the one-hot encoded label using crossentropy function. This entire setting is often called multinomial logistic classifications.



BONUS

5) CNN slide number 23

i) Sum Pooling

$1+1+5+6 \Rightarrow 13$

1	1	2	4
5	6	7	8
3	2	1	0
1	2	3	4

$\Rightarrow 21$

$\Rightarrow 8$

8

$\Downarrow$

13	21
8	8

ii) Average Pooling

$\frac{13}{4}$ $\frac{\text{sum}}{\text{number}}$

3.25

1	1	2	4
5	6	7	8
3	2	1	0
1	2	3	4

$\Rightarrow 21/4 = 5.25$

$\Rightarrow 8/4 = 2$

$\frac{8}{4}$

2

3.25	5.25
2	2