Bo Lin 1001778270

1.

(a) T (b) T (c) T (d) F (e) T

(f) F (g) T (h) T (i) F (j) F

(k) T (l) T (m) T (n) F (o) F

2.

(a) Boolean search often result in either too few or too many. Feast means too many and famin means too few.

(b)

Overfitting is the phenomenon of fitting a particuler dataset too closely or precisely to fit other data well or to predict future observations.

(c) Auto encoder is an artificial neural network that is used for efficient encoding in unsupervised learning. The purpose of auto encoder is to learn a representation of a set of data. usually for dimension reduction.

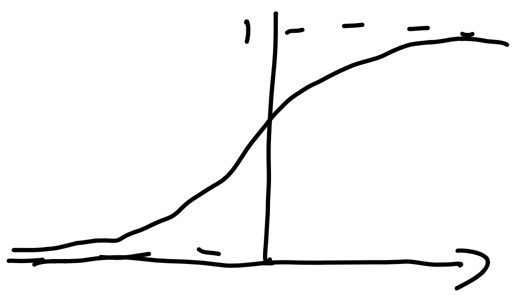
(d) sigmoid is a S shaped curve

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

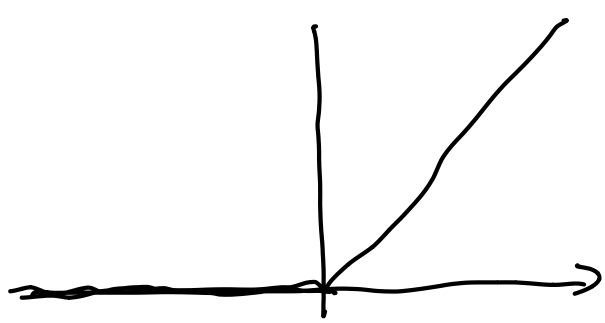
ReLU is a activation function which has less nonlinear then sigmoid function

the formula is $ReLU(x) = \max(0, x)$

sigmoid



ReLU



3.

(a).

$$\text{Jaccard coef} = \frac{|\text{Batman} \cap \text{Robin}|}{|\text{Batman} \cup \text{Robin}|}$$

$$= \frac{4}{5}$$

(b)

after processing:

S1 : Dr Kim graduated University Wisconsin Madison

S2 : Prof kim now works University Texas Arlington

$$\text{Jaccard coef} = \frac{|S_1 \cup S_2|}{|S_1 \cap S_2|}$$

$$= \frac{2}{11}$$

4.

(a)

$$\text{likelihood}(\mu, \sigma) = p(x|\mu, \sigma^2)$$
$$= \prod_{i=1}^{N} N(x_i|\mu, \sigma^2)$$

(b)

$$L(\mu, \sigma) = p(x|\mu, \sigma^2)$$
$$= \prod_{i=1}^{N} N(x_i|\mu, \sigma^2)$$

$$\ln p(\mu, \sigma^2) = \ln \prod_{i=1}^{N} N(x_i|\mu, \sigma^2)$$
$$= \sum_{i=1}^{N} \ln N(x_i|\mu, \sigma^2)$$
$$= -\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\mu)^2 - \frac{N}{2}\ln\sigma^2 - \frac{N}{2}\ln(2\pi)$$

(c)

$$l(\mu, \sigma) = \ln p(\mu, \sigma^2)$$

$$\mu_{MLE}, \sigma_{MLE} = \text{argmax}_{\mu,\sigma} \ln p(x|\mu, \sigma^2)$$

$$\mu_{MLE} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\sigma^2_{MLE} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_{MLE})^2$$

5.

(a)

$$P(y=1 \mid x_1 \dots x_n) = \frac{P(y=1)\prod_{i=1}^{n} P(x_i \mid y=1)}{P(x_1 \dots x_n)}$$

$$= \frac{P(y=1)\prod P(x_i \mid Y=1)}{P(Y=1)\prod P(x_i \mid Y=1) + P(Y=0)\prod P(x_i \mid Y=0)}$$

$$= \frac{1}{1 + \frac{P(Y=0)\prod P(x_i \mid Y=0)}{P(Y=1)\prod P(x_i \mid Y=1)}}$$

$$= \frac{1}{1+\exp\left(\ln\left(\frac{P(Y=0)\prod P(x_i \mid Y=0)}{P(Y=1)\prod P(x_i \mid Y=1)}\right)\right)}$$

$$= \frac{1}{1+\exp\left(-\ln\left(\frac{P(Y=1)\prod P(X_i \mid Y=1)}{P(Y=0)\prod P(x_i \mid Y=0)}\right)\right)}$$

$$= \frac{1}{1+\exp\left(-\ln\left(\frac{P(Y=1)}{P(Y=0)}\right) - \sum_{i=1}^{n} \ln\left(\frac{P(x_i \mid Y=1)}{P(x_i \mid Y=0)}\right)\right)}$$

So. $W_0 = \ln\left(\frac{P(Y=1)}{P(Y=0)}\right)$

$w_i x_i = \ln\left(\frac{P(Y=1)}{P(Y=0)}\right)$

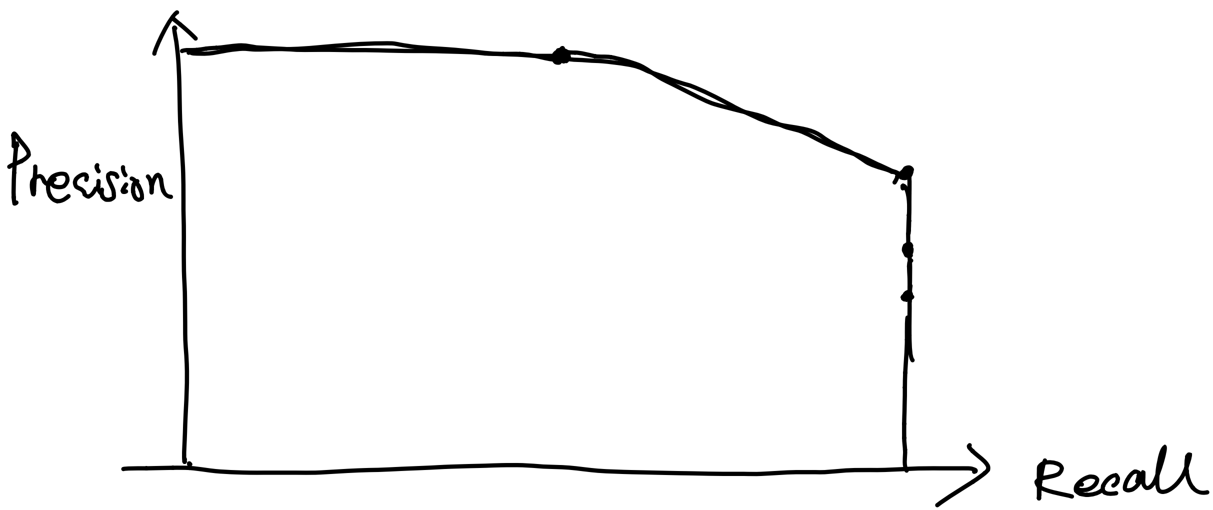
(b). Naive Bayes and logistic Regression have same function form. and have same hypothesis space bias

Naive Bayes and logistic Regression use different method to estimate Naive Bayes is a generative method and logistic regression is a discriminative method

6

(a)

threshold	TP	FN	FP	TN	Precision	Recall
0.1	2	0	3	0	0.4	1
0.12	2	0	2	1	0.5	1
0.4	2	0	1	2	0.667	1
0.8	1	1	1	2	1	0.5
0.99	0	2	1	2	0	0



threshold 0.1

	P	N
P	2	3
N	0	0

threshold 0.12

	P	N
P	2	2
N	0	1

threshold 0.4

	P	N
P	2	1
N	0	2

threshold 0.8

	P	N
P	1	1
N	1	2

threshold 0.99

	P	N
P	0	1
N	2	2

7

a. $$z^{(d)} = w_0 + \sum_1^n w_i x_i^{(d)}$$

$$out^{(d)} = \frac{1}{1+e^{-z(x)}}$$

$$E(x) = \sum_i - y_i \ln(O_i) - (1-y_i)\ln(1-O_i)$$

$$\frac{\partial o^{(d)}}{\partial z^{(d)}} = o^{(d)}(1-o^{(d)})$$

$$\frac{\partial E^{(d)}}{\partial w_i} = \frac{\partial E^{(d)}}{\partial o^{(d)}} \frac{\partial o^{(d)}}{\partial z^{(d)}} \frac{\partial z^{(d)}}{\partial w_i}$$

$$\frac{\partial E^{(d)}}{\partial o^{(d)}} = \frac{o^{(d)} - y^{(d)}}{o^{(d)}(1-o^{(d)})}$$

$$\frac{\partial o^{(d)}}{\partial z^{(d)}} = o^{(d)}(1-o^{(d)})$$

$$\frac{\partial z^{(d)}}{\partial w_i} = x_i^{(d)}$$

$$\frac{\partial E}{\partial w_i} = (o^{(d)} - y^{(d)})x_i^{(d)}$$

(b) $O = \text{sigmoid}(wx + b)$

$= \frac{1}{1 + e^{-(1.5)}}$

$= 0.8176$

$Z(w) = -\ln(0.8176)$

$\frac{\partial Z}{\partial Z_j} = y_i - O_i$