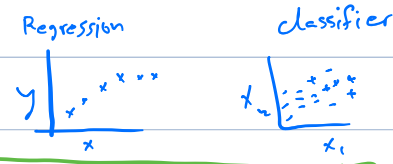
Linear classifier ⇒ Linear decision baundary

may use features expantion for complicated distribution



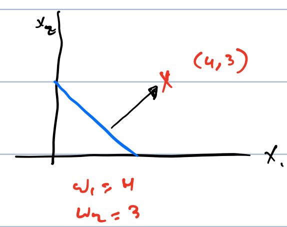
★ Perceptron:-

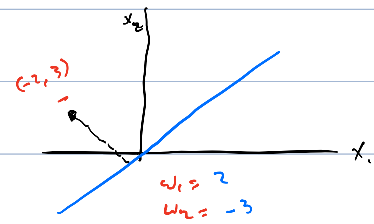
⇒ Assumption ⇒ there is exist Linear classifier
if not ⇒ infinit Loop

$x_1w_1 + x_2w_2 + b = 0$
↓
Shifter

it provides a random seperator

the best line to classify
when it's ⊥ to the sample





relative
(two classes)

+ve Sample ⇒ $W = W + X$

-ve Sample ⇒ $W = W - X$ ⇒ b will be including in W and it's x will be 1

→ Start with $w_1 = w_2 = 0$

$y(x_1w_1 + x_2w_2) > 0$ ⇒ correct classify

if $y(x_1w_1 + w_2w_2) \le 0$.

if y is +ve:-

$w_1 = w_1 + x_1$
$w_2 = w_2 + x_2$

elseif y is -ve:-

$w_1 = w_1 - x_1$
$w_2 = w_2 - x_2$

⟺

if $y^i(\sum w_j x_j) > 0$

else if $y^i(\sum w_j x_j) \le 0$

$\vec{w} \leftarrow \vec{w} + y^{(i)} x^{(i)}$

$b \leftarrow b + y^i$ -1 or 1

→ repeat untill all samples correctly classified classification Rule = $y_i = \text{Sign}(\Sigma wx)$

* \# of iteration that will take until find a solution $= \left(\frac{R}{\gamma}\right)^2$ ↗ longest, ↘ closest

* the solution of Linear classifier depending on the order of samples

---

-It will give us the result with a confidence rate (how much we are sure that our answer is correct)

-It will solve the problem of running into an infinity loop when we don't have a linear line that separates the data.

affected by Imbalanced data

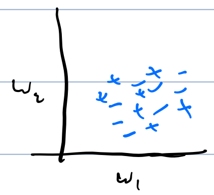
base % = 50% (random)

goal % = human %

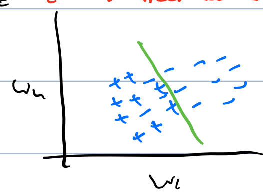
LOGISTIC REGRESSION ⇒ It is a linear classifier

hyper parameters :- learning-rate (if you Regularize)

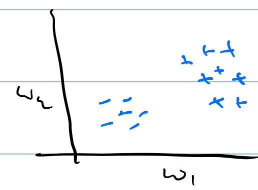
Step-size (doesn't affect accuracy)



there is problem with W



not sure with class but it's good for some fields



Linear classifier

→ farther from decision boundary ⇒ more confidence that right class

logistic function

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

$z = \Sigma w_i x_i + b$

0 → 0.5 → from positive side

+ve → > 0.5

+∞ → 1

-ve → < 0.5

-∞ → 0

testing

Training :-

$$\prod_i P(y^i = +1/- \mid w, x, b)$$ → should be maximaize to get good classifier

$$-\sum_i \log P(y^i = +/- \mid w, x, b)$$

$\hat{y} = \sigma(wx+b)$

Loss and total cost:-

$$L(y,\hat{y}) = -y \log \hat{y} - (1-y) \log(1-\hat{y})$$

$$J = \frac{1}{m} \sum -y \log \hat{y} - (1-y) \log(1-\hat{y})$$

Cross-entropy

* there is no closed-form solution because there is no inverse of the fun.

* gradient descent:- (same as linear regression but with new dw and db)

1 → $w, b = 0$

2 → J

3 → $\frac{dJ}{dw}, \frac{dJ}{db}$

4 → $w_j \leftarrow w_j - \frac{dJ}{dw} \rightarrow \frac{1}{m} \sum (\hat{y} - y) x_j$

$b \leftarrow b - \frac{dJ}{db} \rightarrow \frac{1}{m} \sum (\hat{y} - y)$

5 → repeat 2 - 4 until convergence

$J^{k+1} - J^{k} \leq (10^{-6})$

$\frac{dL}{dw} = (\hat{y} - y) x$, $\hat{y} = \sigma(wx+b)$

$\frac{dL}{db} = (\hat{y} - y)$

* Regularization is very important for Logistics Regression

① Ridge

② Lasso

③ early stopping

* We deal with overfitting and underfitting like linear regression

- Perform training and check the error:
- If error is high: (underfitting)
  - Add more features
  - More complex model by polynomial
- If overfitting:
  - Add more data
  - Perform regularization

* Sentiment analysis

Example of Logistics Regression =>

BOW → bag of words

→ we lost the sequence of words
we can solve it by taking n-gram

→ the size of each sample is big

Preprocessing :-

1- punctuation removal

2- Stop word removal

3- case normalization

4- remove root word

The length of sample will be same for all samples, and it will be vector of what word is there or not

* margin :-

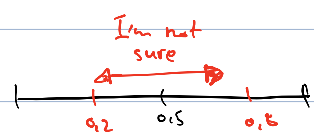
In the binary classification problem, if we want to use linear regression, we will give a threshold value and the probability that higher than this will be positive, otherwise negative. Normally, it is 50%:

margin = 0.3

if $y \geq 0.5$:
    y = +ve
else:
    y = -ve

→

if $y \geq 0.8$:
    y = +ve
else $y \leq 0.2$:
    y = -ve
else:
    I'm not sure



* having margin = 0.45, it doesn't mean you are 95% confident

* to get the most positive and most negative words :-

Sort the W vector. then take the Highest and Lowest words

because word with most positive has most maguited of w

Multiclass

logitic regression for every class

* SoftMax $\dfrac{e^{z_i}}{\sum_j^n e^{z_j}}$, $z_i = wx + b$

do one vs All classification ⇒ if you have 5 classes: you will get 5 classifiers

Imbalence data sets:- accurecy is not a good indecator for it

when a class has much more data than another

* normally:- the class with fewer data is considered +ve

Recall and Precision

not good with Imbalance DS

		Actual +ve	Actual -ve
Predict	+ve	TP	FP
	-ve	FN	TN

Percision

recall

$$\text{accurecy} = \frac{TP + TN}{\text{all}}$$

الإحكام

$$\text{precision} = \frac{TP}{TP + FP} \qquad \text{recall} = \frac{TP}{TP + FN}$$

for regression: MSE, ASE

Precision goes opposite recall

F-B Score

combine P and R

$$F_B = (1 + B^2) \frac{P \times R}{B^2 P + R}$$

$B \in [0, 2]$

calculate the mean with different wiegts without Prefering anyone

$$F_1 = 2\,\frac{P \times R}{P + R}$$

$\beta = [0,1) \Rightarrow$ Prefer Recall

$\beta = 1 \Rightarrow$ without Prefering

$\beta = (1,2] \Rightarrow$ Prefer Precision

※ P and R for multiclass

using confusion matrix

		Actual A	Actual B	Actual C
Prediction	A	10	2	1
	B	1	7	2
	C	3	2	15

TP

$P_B = \frac{7}{7+2+1}$ $\quad P_C = \frac{15}{15+2+3}$

$R_B = \frac{7}{7+2+2}$ $\quad R_C = \frac{15}{15+2+1}$

※ How to deal with Imbalance Data

# of Samples	
class A	class B
180	20

① Data Level → on train data

① under sampling → select 20 random samples from class A

② over sampling → ① make copies of class B to get 180 sample

③ hyped ① and ② — with numeric data — ② SMOTE ⇒ create new samples between samples

sleep	study	Grade
6	6	Pass
5	6	Pass
5.5	6	Pass

new sample → (5.5, 6, Pass)

④ change the weight of class in cost function

$$J = \frac{1}{m} \sum -\log y \log \hat{y} - {}^{b}(1-y) \log(1-\hat{y})$$

※ How to deal with categorical features:-

One-hot encoding → * avoid wieghted features
- adding more features
- will not work if you have 2 anemals in one image

Empeding

* Zip-code should be categorical to avoid weight