21-9-2021 Tuesday

week # 4 lecture # 5

Linear classifier:

test time matters more to us than Training time

Cosine similarity

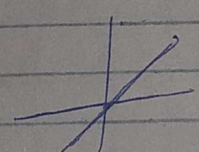
$(a-b)^2 = a^2+b^2-2ab$

$= ab$

Initially, we initialize w with random values & update these values overtime

$y = mx + c$

if $c = 0$,



Line passes through origin

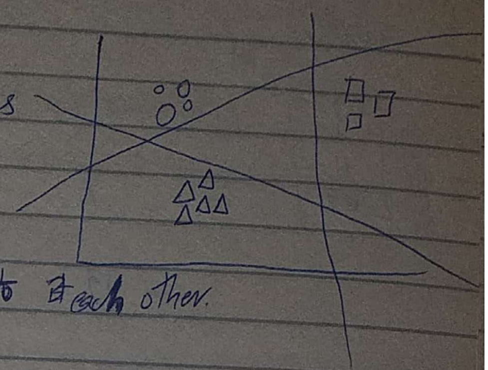
x_1, x_2

$m_1 x_1 + m_2 x_2 + c$

we try to learn a linear line for each class

A Limitations of linear classifier,

it doesn't give good results as it doesn't really differentiate classes if data is too close to each other.



① True class score should be max when training

② True class score should be significantly greater than other classes ~~so difference between classes should also be great~~

loss function

$$\begin{array}{c c} S_y & \\ S_2 & \\ S_3 & \end{array}\begin{bmatrix} 10 \\ 3 \\ 4 \end{bmatrix} \leftarrow \text{cat}$$

$$\max(0, \underbrace{3-10}_{S_2 - S_y} \overset{\text{margin}}{+5}) + \max(0, \underbrace{4-10}_{S_3 - S_y} +5)$$

$$0 + 0$$

$$= 0$$

margin → how much True class score should be greater than other classes

$$\begin{array}{c} S_y \\ S_2 \\ S_3 \end{array}\begin{bmatrix} 8 \\ 3 \\ 4 \end{bmatrix} \quad \max(0, 3-8+5) + \max(0, 4-8+5)$$

$$= 0 + 1$$

$= 1$ loss score

E

for example:

$$\begin{array}{c} S_y \\ S_2 \\ S_3 \end{array} \begin{bmatrix} 8 \\ 11 \\ 7 \end{bmatrix} \text{cat} \quad L(x) = \max(0, 11-8+\boxed{5}) + \max(0, 7-8+5)$$

margin/hyperparameter

$$= 8 + 4$$

$$= \underline{12}$$

$$\begin{array}{c} S_y \\ S_2 \\ S_3 \end{array} \begin{bmatrix} 10 \\ 2 \\ 2 \end{bmatrix}$$

Loss function for one ~~data~~ calculation

Cost " for whole train set

$$\text{Cost function} = \frac{1}{m} \sum_{i=1}^{m} L_i(x_i)$$

$\therefore$ m = ~~no~~ no. of Training imgs/example

hinge loss function

$\Delta = \cancel{8}5$ = margin

$$\boxed{L(x) = \sum_{j \neq y} \max(0, S_j - S_y + \Delta)}$$

$\Delta = 1$ usually used

SVM = Support ~~vector~~ vector Machine classifier we also add a bias in the Wx matrix

$$\cancel{\begin{bmatrix}3\\3\\4\\5\end{bmatrix}}$$

for

$$\left[\begin{array}{cccc|c} 1 & 2 & 3 & 4 & 2 \\ 1 & 1 & 2 & 3 & 3 \\ 4 & 4 & 5 & 5 & 4 \end{array}\right] \begin{bmatrix}2\\3\\1\\2\\1\end{bmatrix}$$

bias added in W

1 Appended

Now To Solve Regression Problem from linear classifier.

Linear Regression..

we change dimensions of W

$$\underset{1\times4}{\begin{bmatrix}8 & 4 & 5 & 2\end{bmatrix}} \underset{4\times1}{\begin{bmatrix}2\\2\\3\\4\end{bmatrix}} = [5.8]$$

Predicted value $f(x)$

Loss function = L_2 distance =

MSE $\frac{1}{2m}\sum (f(x_i) - y_i)^2$

e.g. $(5.8-6)^2$

$= 0.04$ → Closer to Zero means It is good

$$\frac{1}{m}\sum_{i=1}^{m}\left(\sum_{j\neq y}\max\left(0, S_j - S_y + 1\right)\right) + \sum_{i=1}^{c}\sum_{j=1}^{d} w_{ij}^2$$

L_2 Regularization loss
c = classes
d = dimensions

$$\text{True}\begin{bmatrix}20\\15\\201\end{bmatrix} \quad \begin{bmatrix}20\\15\\22\end{bmatrix} \quad \begin{bmatrix}20\\15\\102\end{bmatrix}$$

we need some regularization for our Loss function.

when weights is zero, then regularization is zero
To modify or better the weights we use this regularization loss

$$\underset{w}{\begin{bmatrix}10\\20\\21\end{bmatrix}} \qquad \underset{2w}{2\begin{bmatrix}20\\40\\42\end{bmatrix}}$$

This is better, faster

Small w is preferential

$$\begin{bmatrix}1 & 1 & 1 & 2\\3 & 5 & 6 & 1\\2 & 3 & 4 & 2\end{bmatrix}$$

$c = 3$ 3 classes
$d = 2\times 2 = 4$ Picture was 2×2

$$w_{ij}^2 = 1^2+1^2+1^2+2^2+3^2+5^2+6^2+1^2+2^2+3^2+4^2+2^2$$

example

$\begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix}$ x

$\begin{bmatrix}1 & 0 & 0 & 0\end{bmatrix}$ W_1 not good bcz some dimensions are not evident at all

$\begin{bmatrix}0.25 & 0.25 & 0.25 & 0.25\end{bmatrix}$ W_2 more good weights y

$$\begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} = 1^2 + 0^2 + 0^2 + 0^2 = 1 = R(w_1)$$

$$\begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 \end{bmatrix} = 4(0.25)^2 = 4 \times \left(\frac{1}{4}\right)^2 = \frac{4}{16}$$

$$R(w_2) = 0.25$$

Regularization loss