Dropout

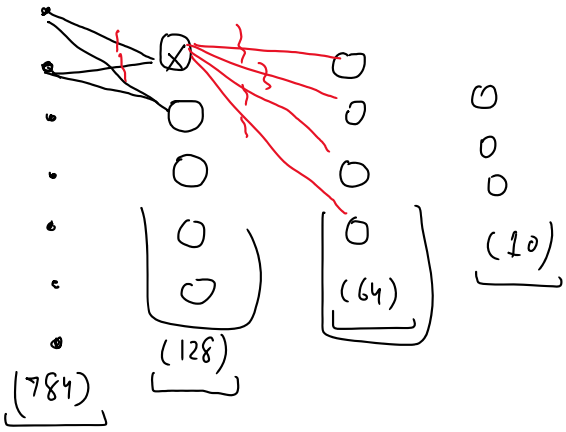
03 January 2025 18:22

[p = 0.5] 50%

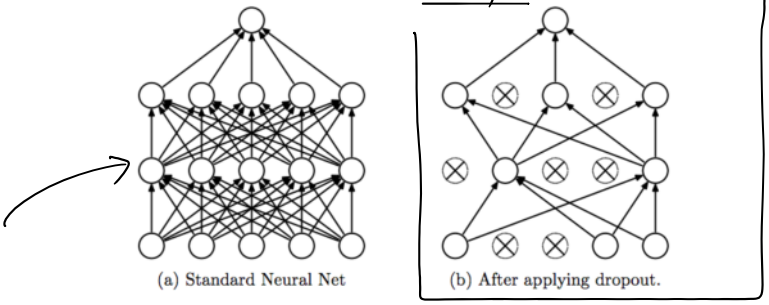
0.3

1. Applied to the hidden layers
2. Applied after the ReLU activation function
3. Randomly turns off p% neurons in the hidden layer during each forward pass
4. This has a regularization effect ✓
5. During evaluation dropout is not used

each forward pass



simplify

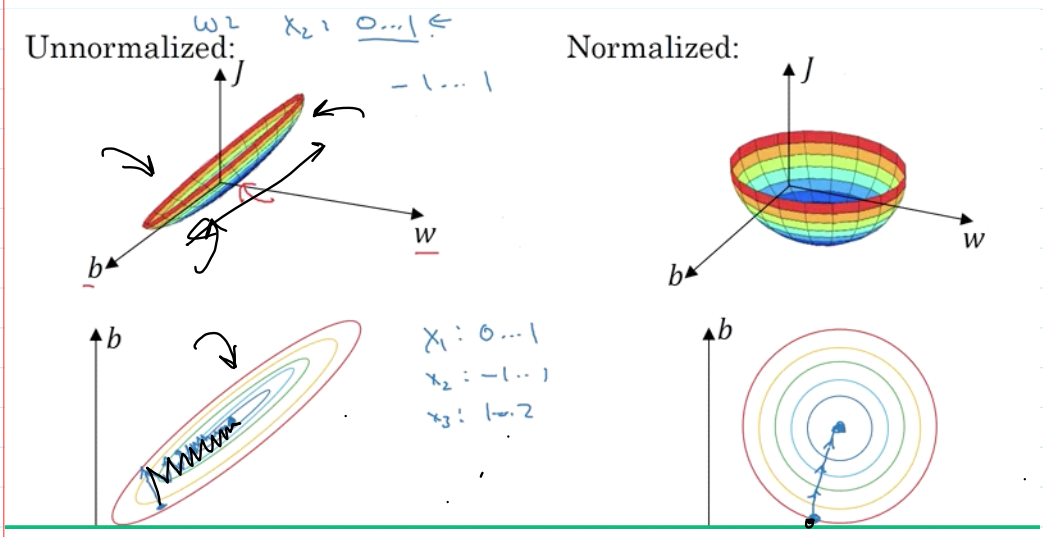


(a) Standard Neural Net

(b) After applying dropout.

Feature Scaling

05 May 2022 17:37



(age) x_1 — x_2 (salary)

w_1 — w_2

$w_2 = w_2 - \eta \frac{\partial L}{\partial w_2}$

Standard

$\frac{x_j - \mu}{\sigma}$

Normalize

$\frac{x_j - x_{min}}{x_{max} - x_{min}}$

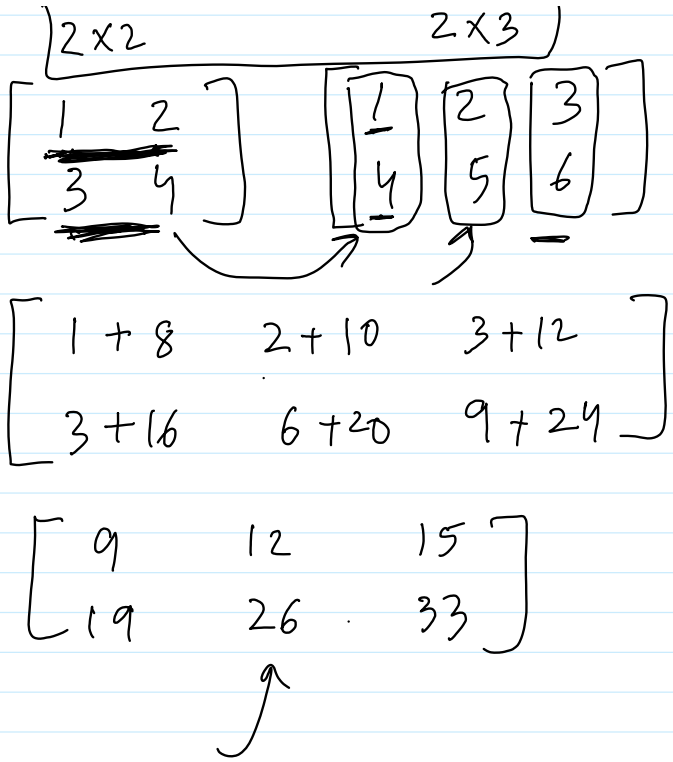
0 to 1

min = 0

max = 1

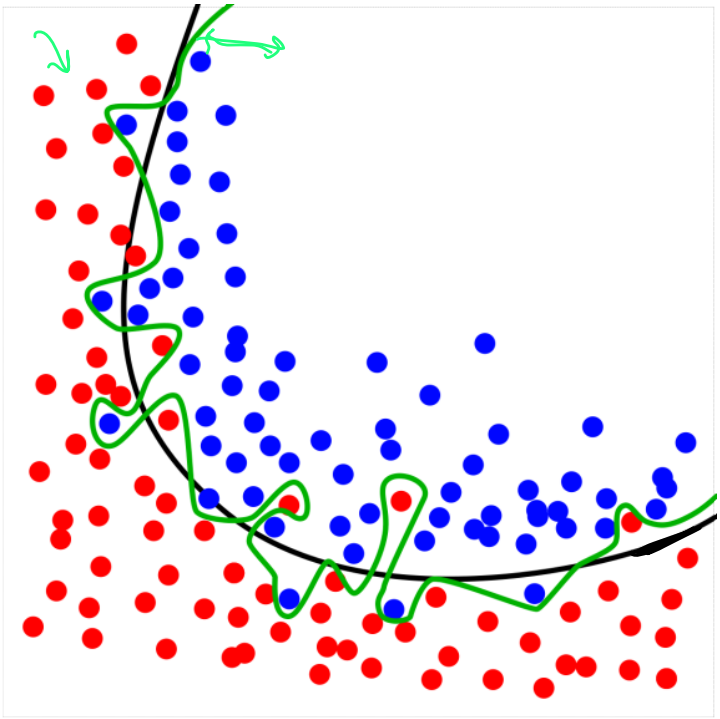
-1 to +1

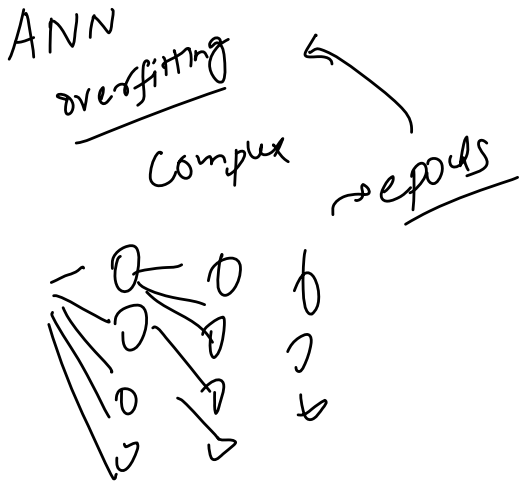
2x2 2x3


2x2
2x3
1 2
3 4
1 2 3
4 5 6
1+8 2+10 3+12
3+16 6+20 9+24
9 12 15
19 26 33

The Problem of Overfitting

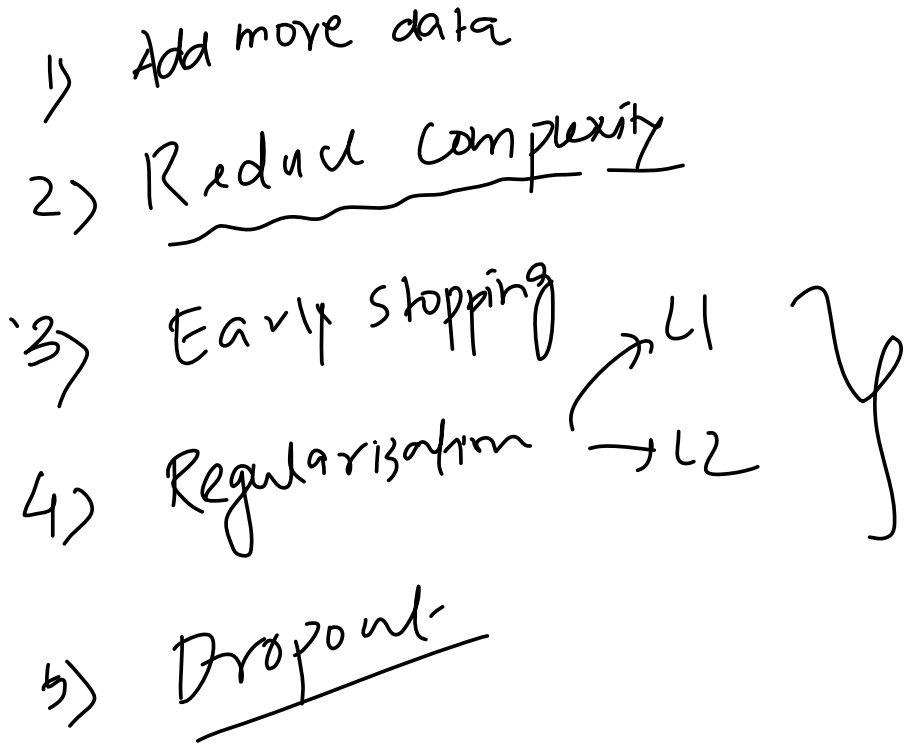
12 May 2022 09:32





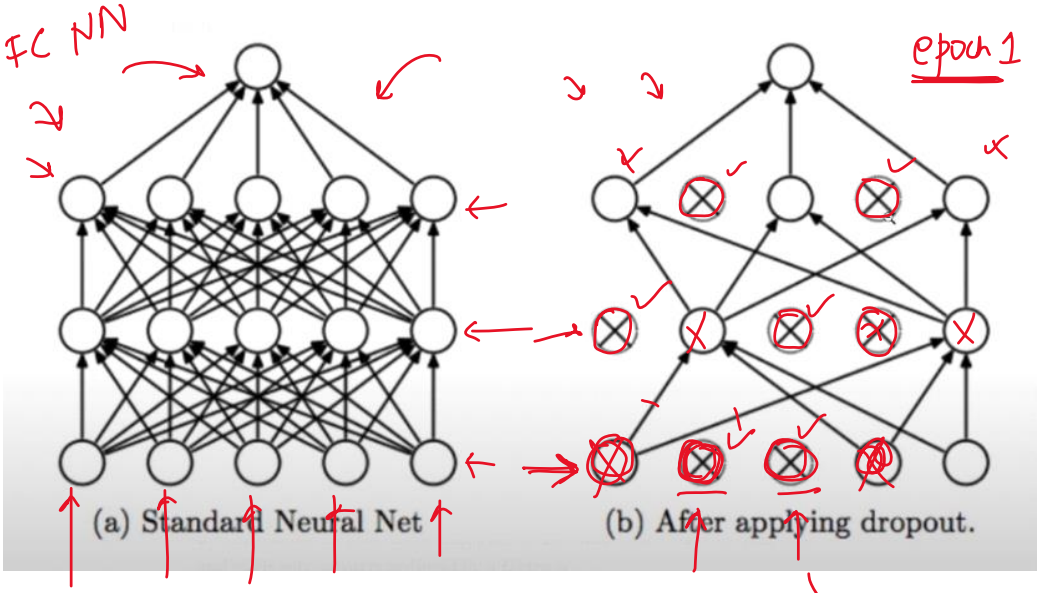
Possible Solutions

12 May 2022 09:33



The concept of Dropouts

12 May 2022 09:33



Dropout

FC NN

epoch 1

$x_1\ x_2\ x_3\ x_4\ x_5\ y$

input nodes

hidden drop

epoch = 10

epoch 2

epoch 3

2·1· → acc

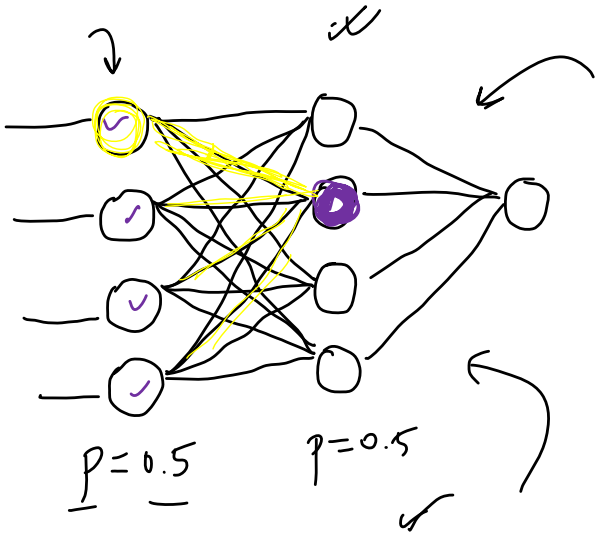
95 → 97

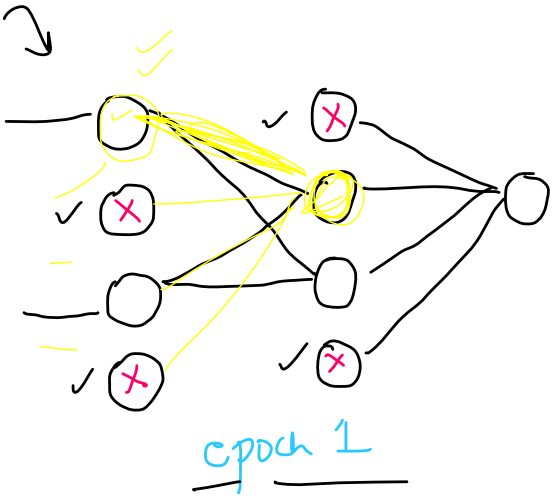
10 epochs

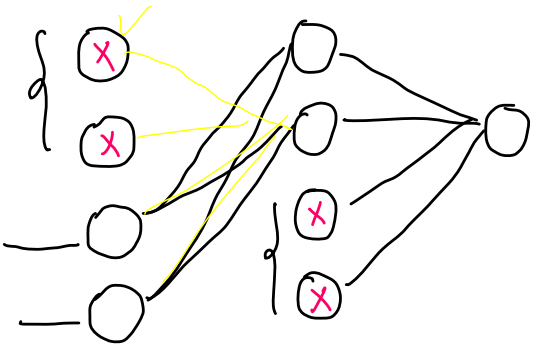
10 diff NNs

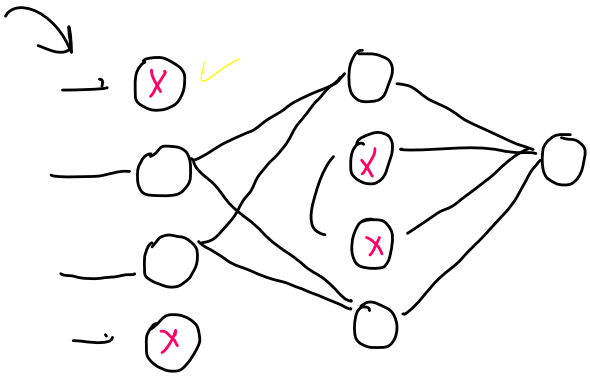
Why this works?

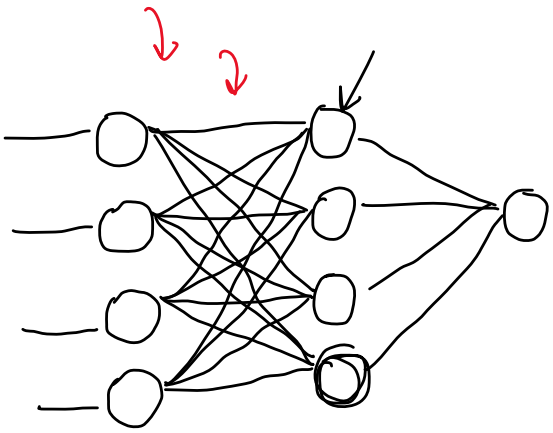
12 May 2022 09:33





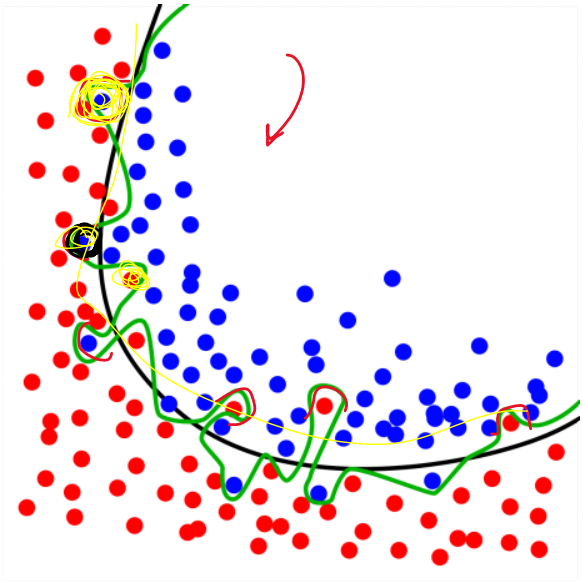






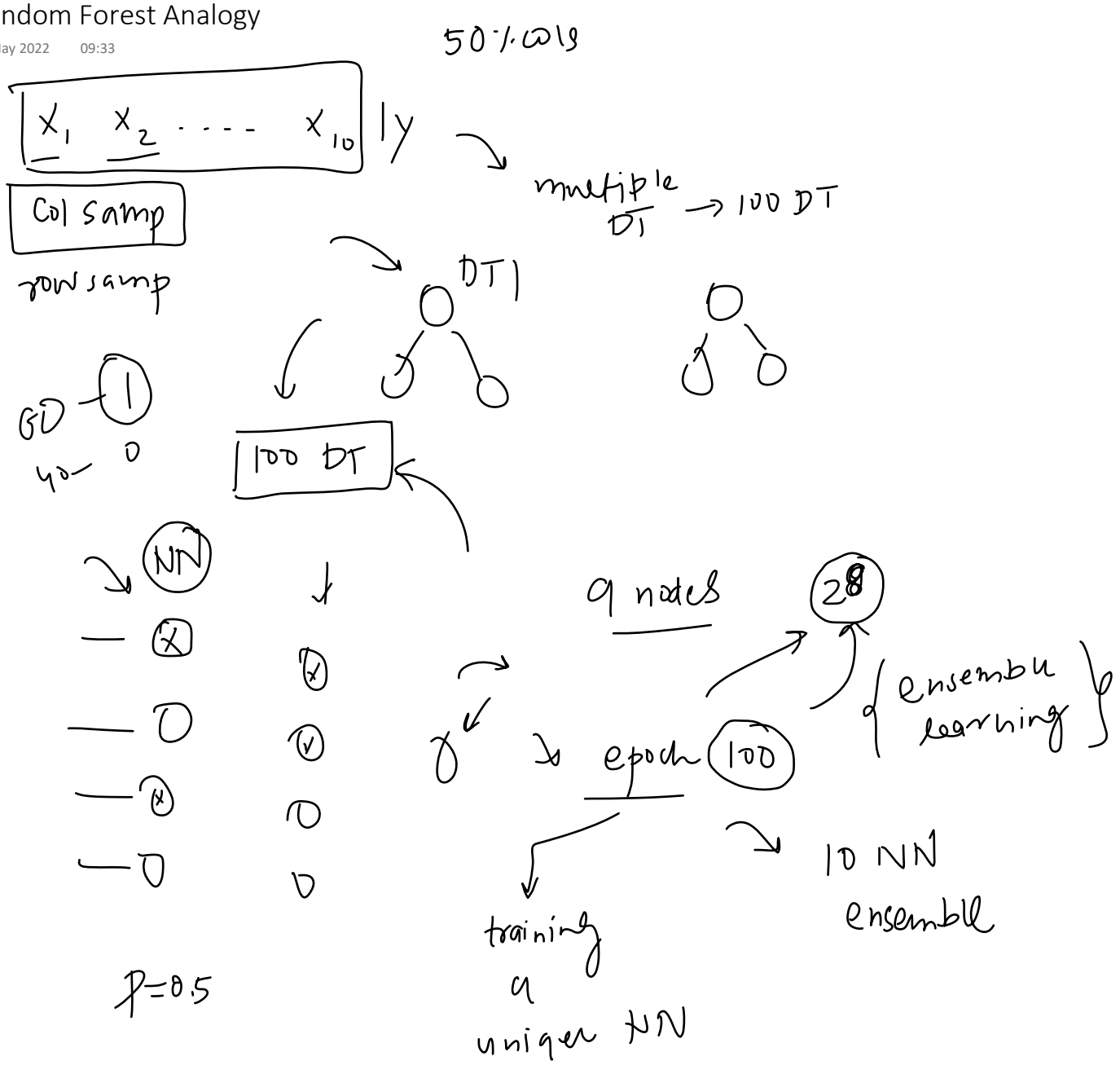
→ Reduce the # nodes

→



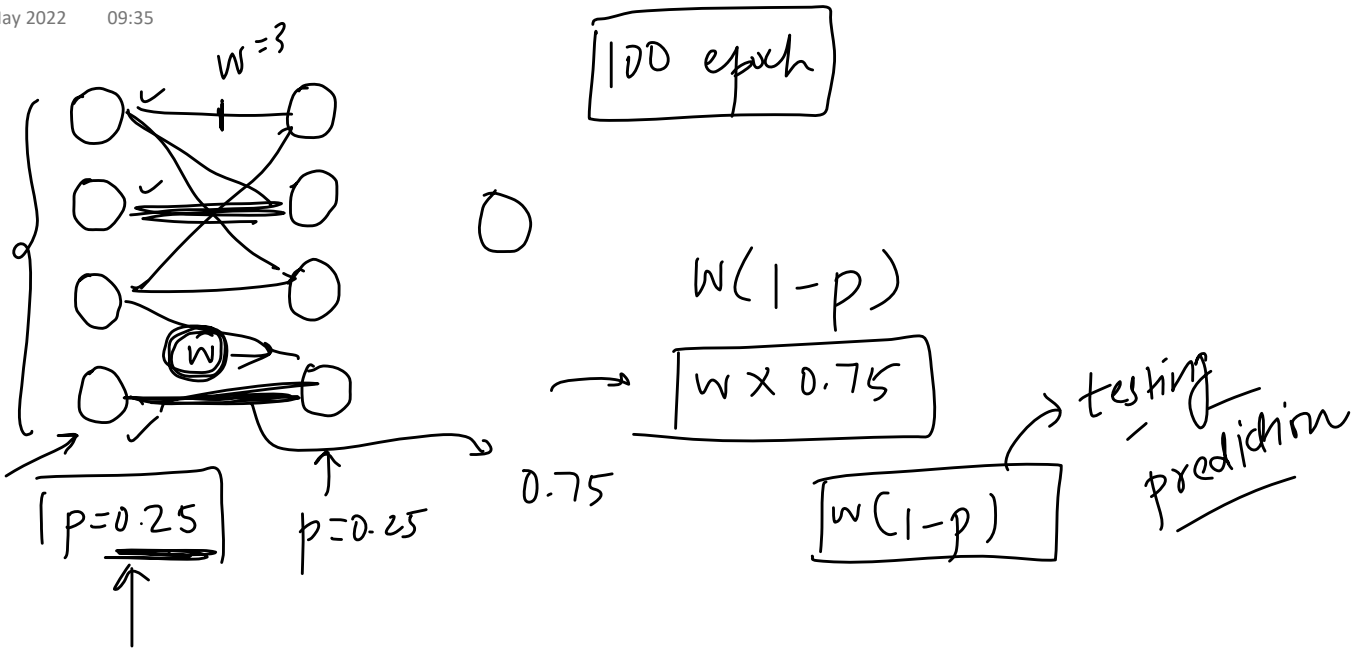
Random Forest Analogy

12 May 2022 09:33



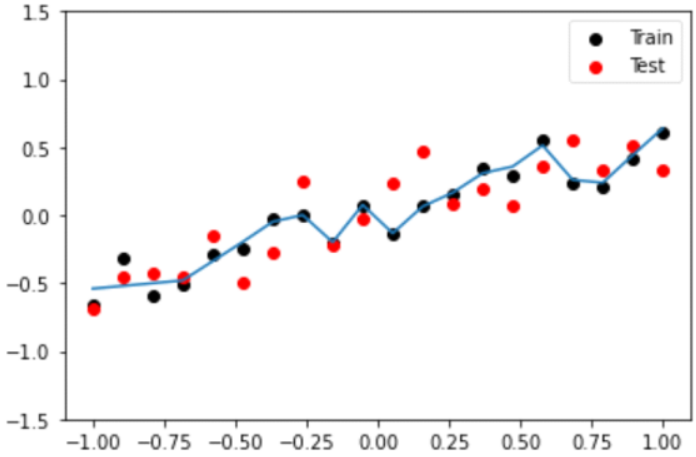
How prediction works?

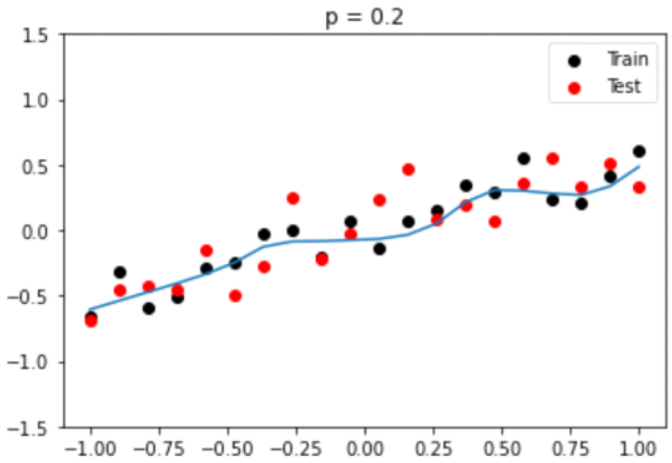
12 May 2022 09:35

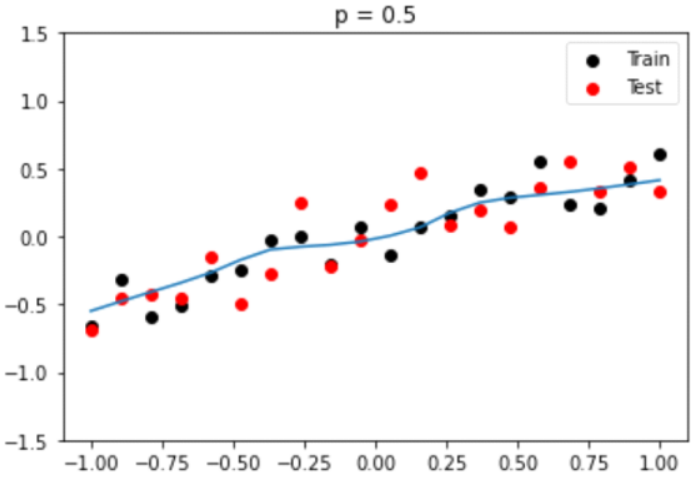


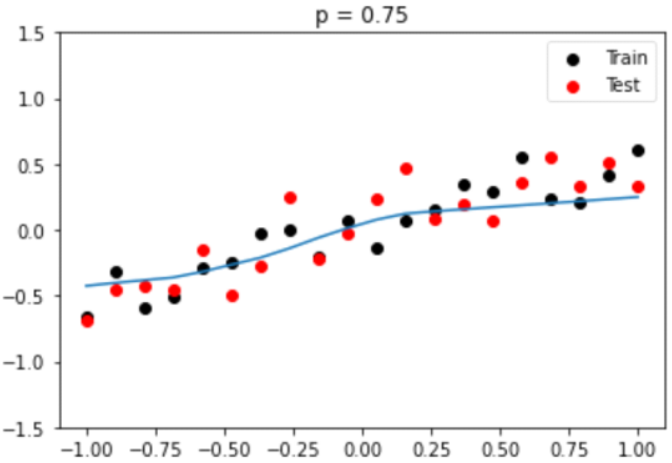
Effect of p

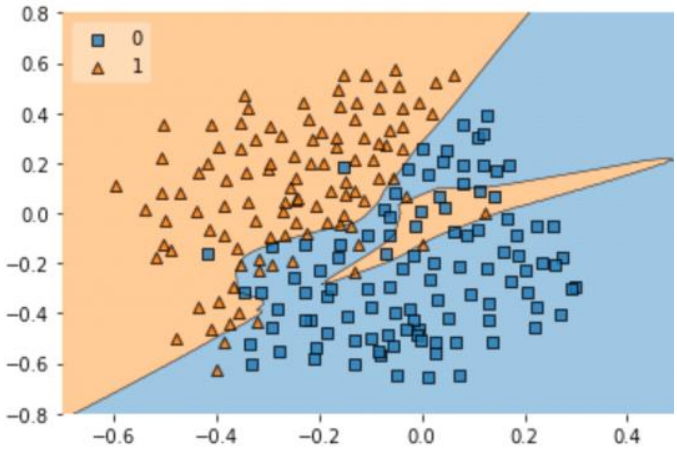
12 May 2022 10:06

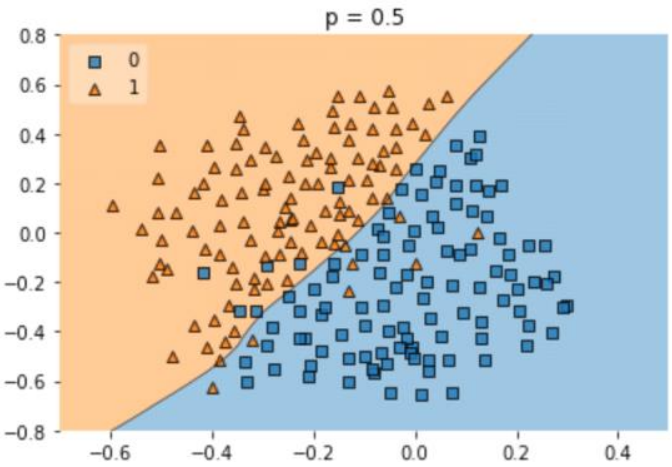










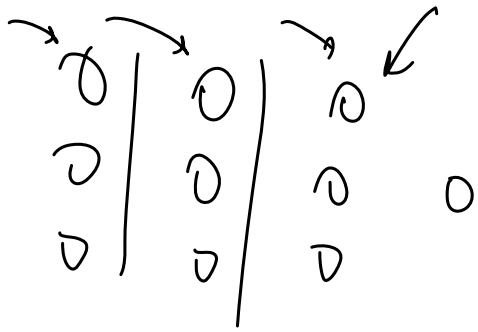


Practical Tips and Tricks

12 May 2022 09:34

1) Overfitting P↑, underfitting P↓

2) Last layer → dropout



3) CNN → 40-50-1 (P) → ✓

20-30 RNN 50>

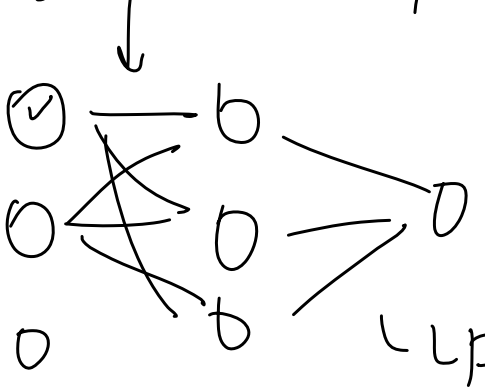
ANN → 10-50

Drawbacks

12 May 2022 09:34

1) convergence → delay

2)



{ loss function } value changes

↳ LF – has to change