

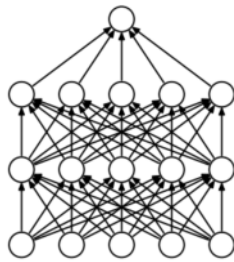
Dropout

03 January 2025 18:22

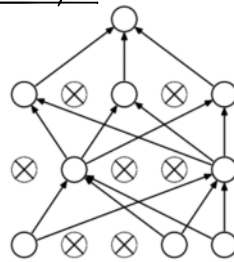
$$[p = 0.5] \quad 50\% \\ 0.3$$

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

simplify

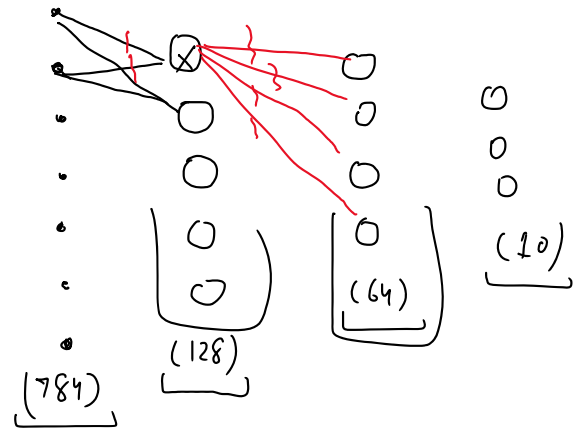


(a) Standard Neural Net

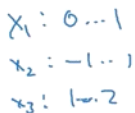


(b) After applying dropout.

each forward pass



05 May 2022 17:37



$X_1 \rightarrow X_2$ (Salary)

W_2

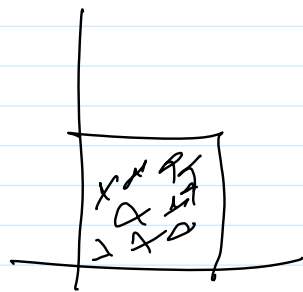
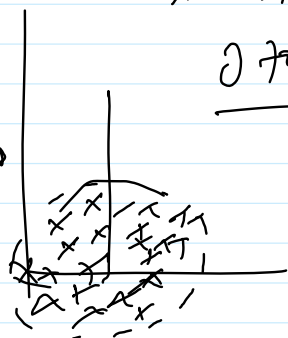
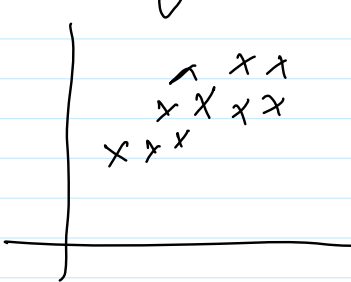
$$w_2 = w_2 - \eta \frac{\partial L}{\partial w_2}$$

Standard

$$\frac{x_j - 14}{s}$$

Normalize

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

$$\min = 0$$

$$-|t_0 + t|$$
$$12 \times 2$$
 2×3

$$\begin{array}{c}
 \begin{array}{|c|c|} \hline 2 \times 2 \\ \hline \end{array} \quad \begin{array}{|c|c|c|} \hline 2 \times 3 \\ \hline \end{array} \\
 \left[\begin{array}{cc} 1 & 2 \\ \hline 3 & 4 \end{array} \right] \quad \left[\begin{array}{|c|} \hline 1 \\ \hline 4 \end{array} \right] \quad \left[\begin{array}{|c|} \hline 2 \\ \hline 5 \end{array} \right] \quad \left[\begin{array}{|c|} \hline 3 \\ \hline 6 \end{array} \right] \\
 \hline
 \end{array}$$

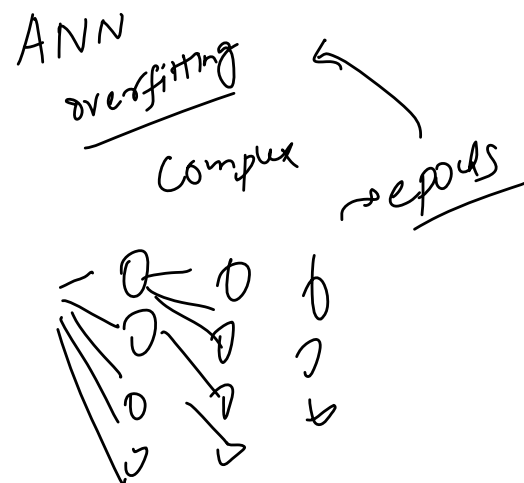
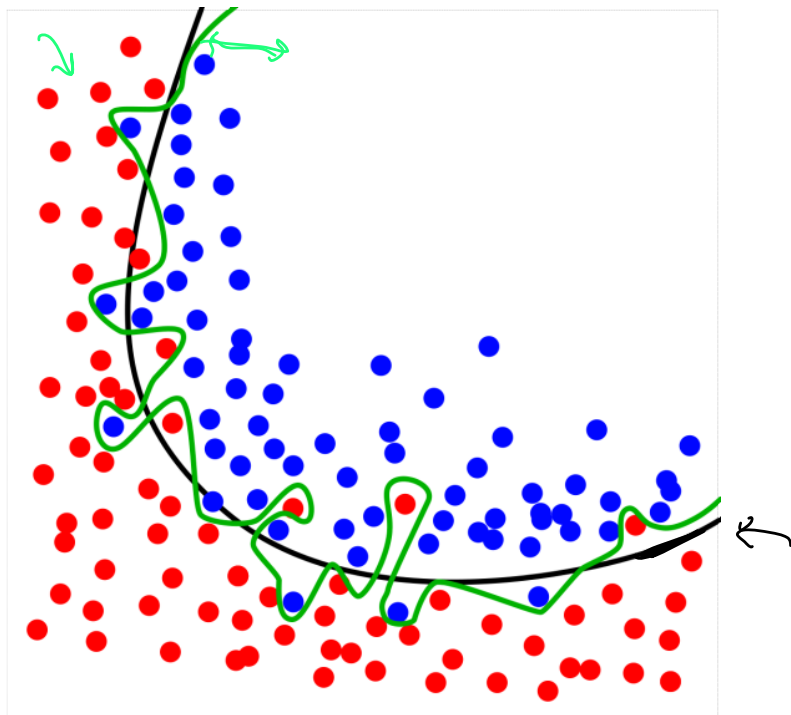
$$\left[\begin{array}{ccc} 1+8 & 2+10 & 3+12 \\ 3+16 & 6+20 & 9+24 \end{array} \right]$$

$$\left[\begin{array}{ccc} 9 & 12 & 15 \\ 19 & 26 & 33 \end{array} \right]$$



The Problem of Overfitting

12 May 2022 09:32



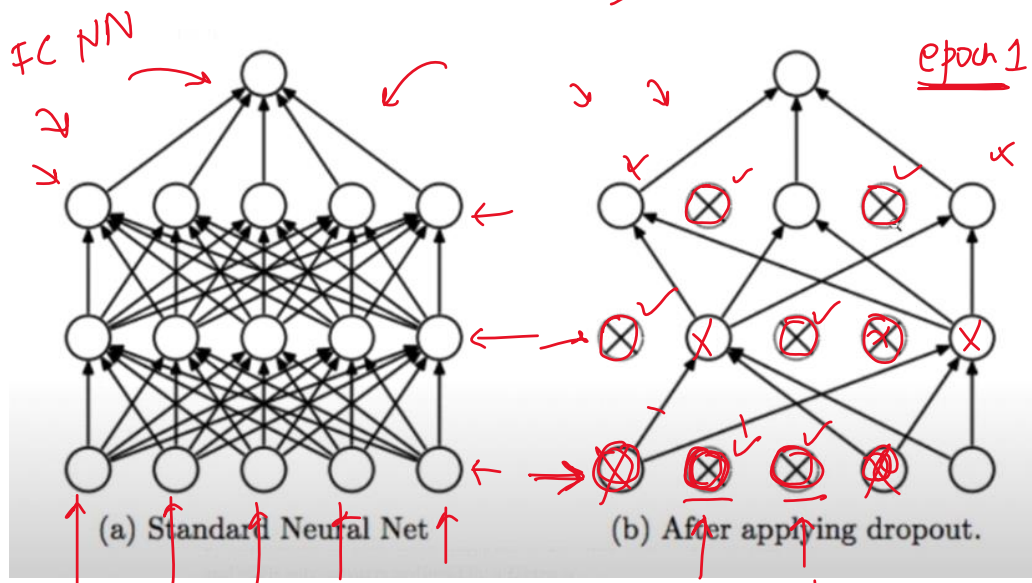
Possible Solutions

12 May 2022 09:33

- 1) Add more data
- 2) Reduce complexity
- 3) Early stopping
- 4) Regularization \rightarrow L1 \rightarrow L2
- 5) Dropout

The concept of Dropouts

12 May 2022 09:33



epoch 2
epoch 3

2.1. → acc

95 → 97

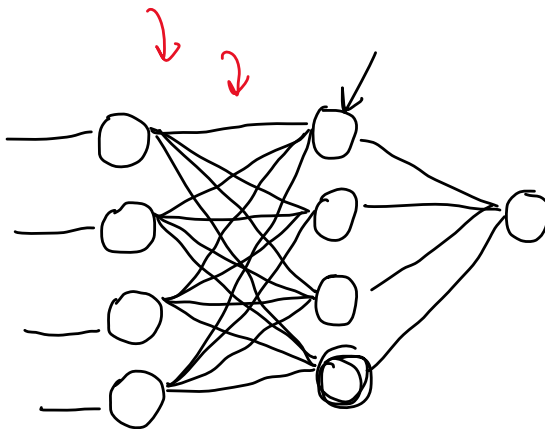
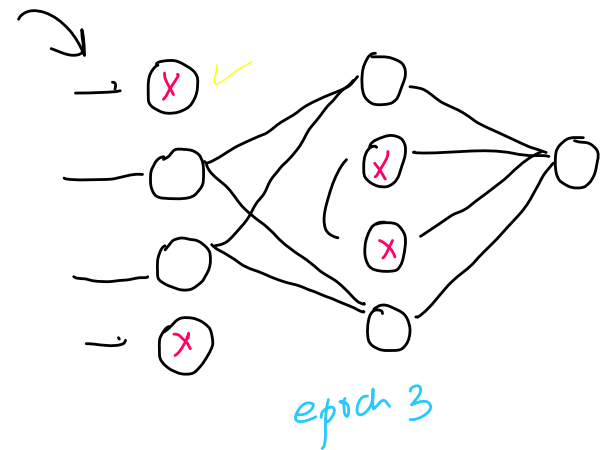
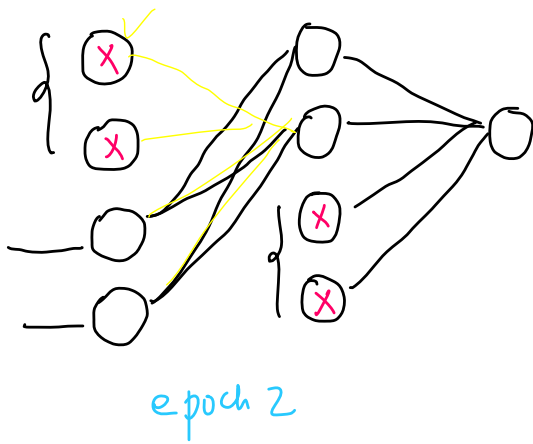
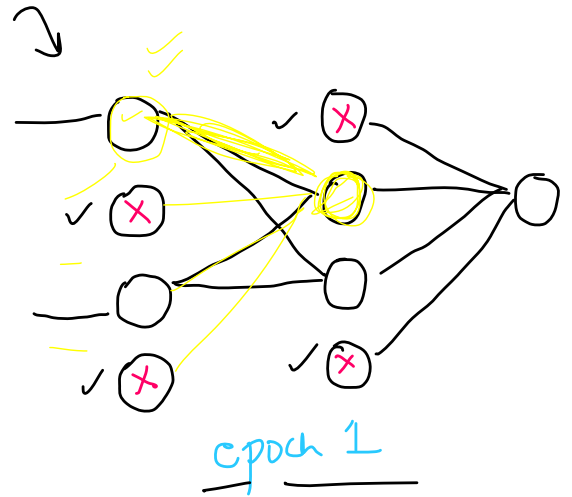
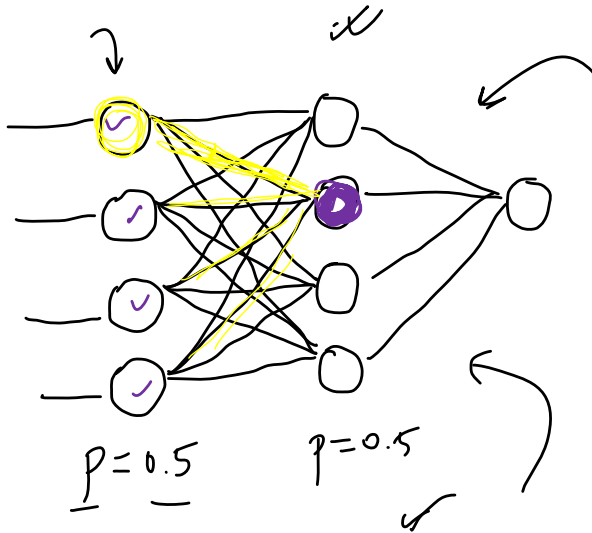
10 epochs

10 diff NNS

The diagram shows the process of training a neural network using dropout. It starts with a standard neural network (a) and then shows the network after applying dropout (b). The diagram is annotated with handwritten notes in red ink, including 'Dropout', 'epoch 1', and 'epoch = 10'. A list of input variables $x_1, x_2, x_3, x_4, x_5, y$ is written to the right. A note 'input nodes hidden drop' points to the input and hidden layers. A box containing 'epoch = 10' is also present. Below the diagram, there are more handwritten notes: 'epoch 2', 'epoch 3', '2.1. → acc', '95 → 97', '10 epochs', and '10 diff NNS'.

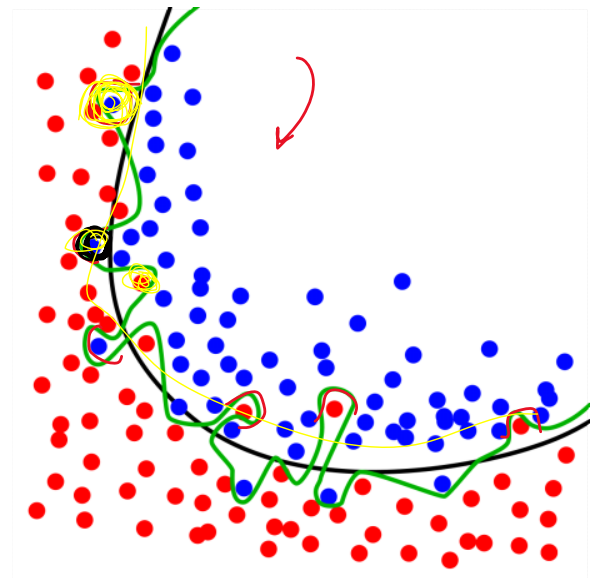
Why this works?

12 May 2022 09:33



→ Reduce the # nodes

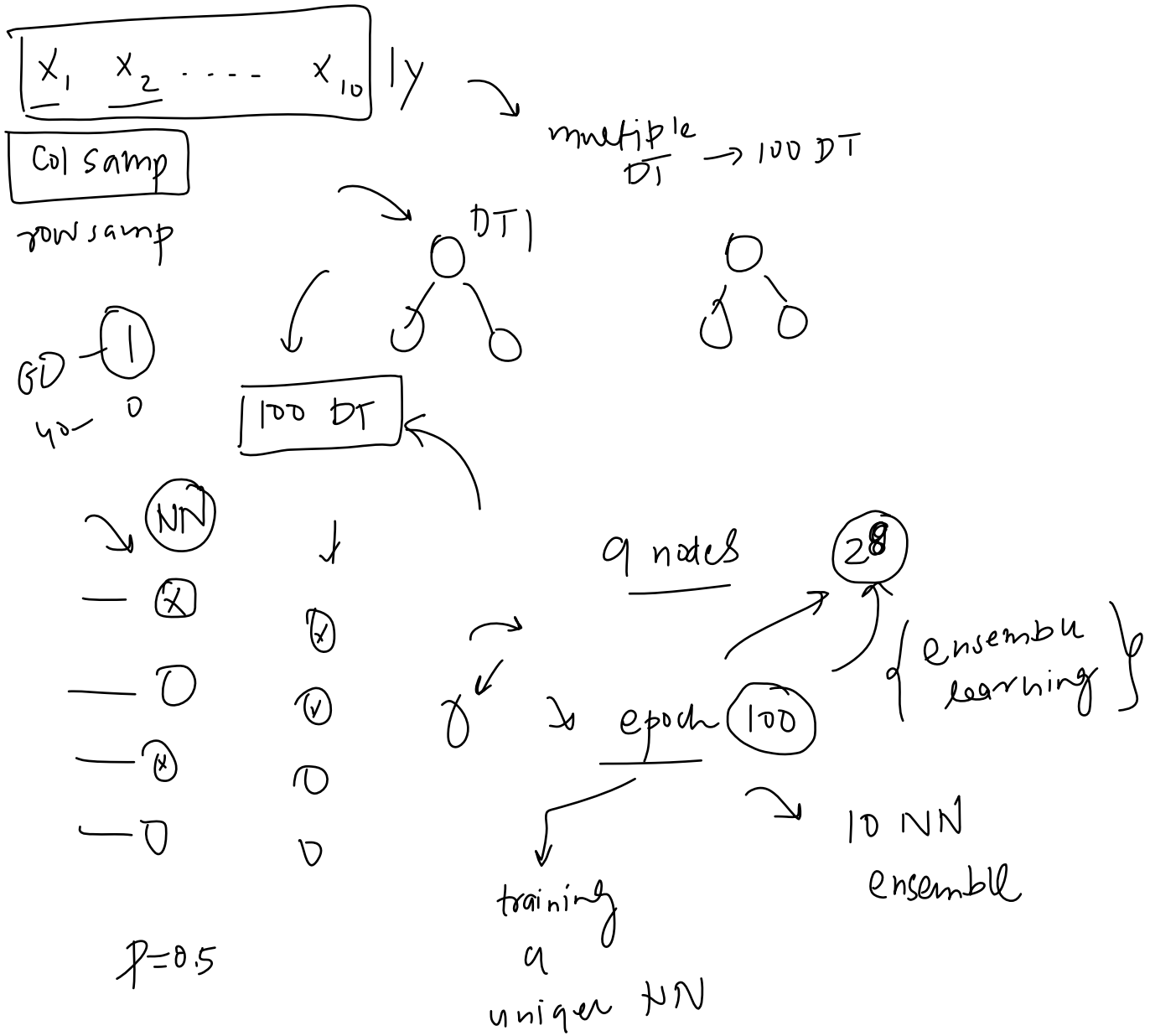
→



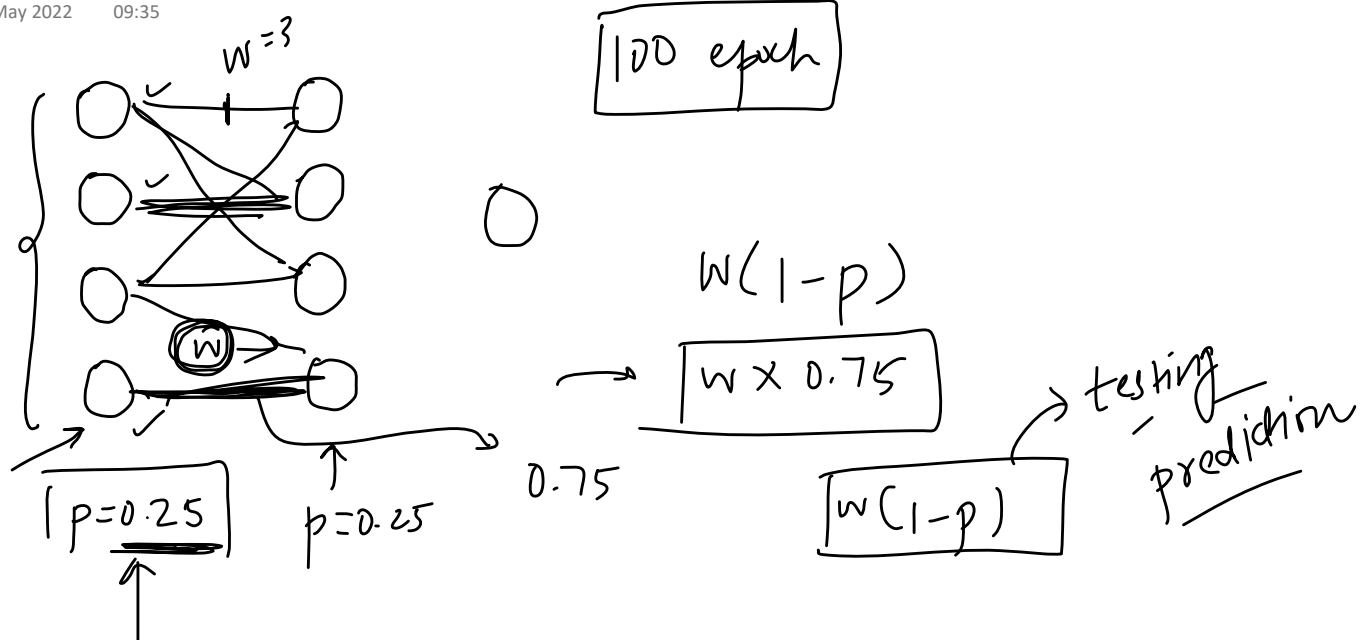
Random Forest Analogy

12 May 2022 09:33

50% w/o



12 May 2022 09:35



Regression Code Example

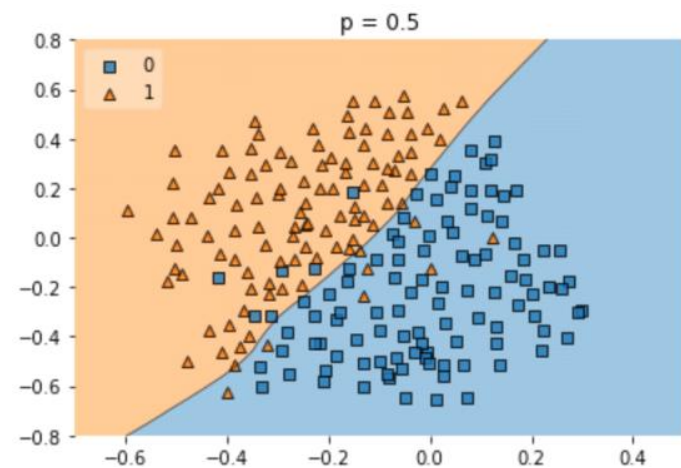
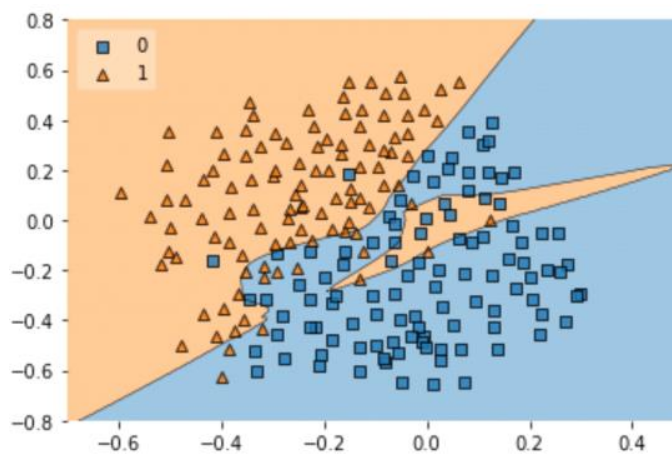
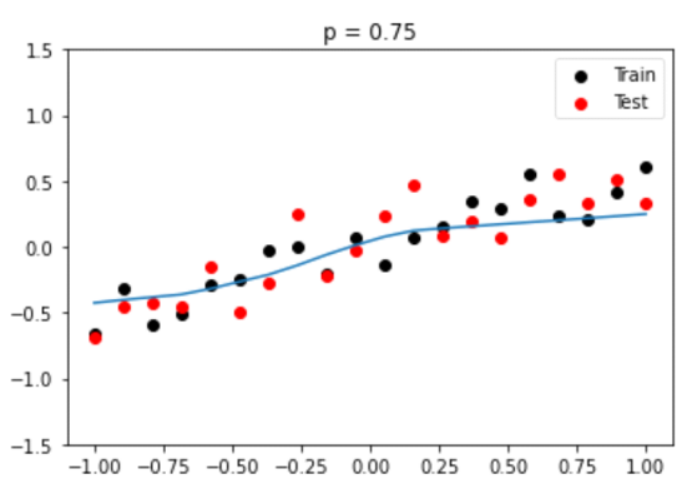
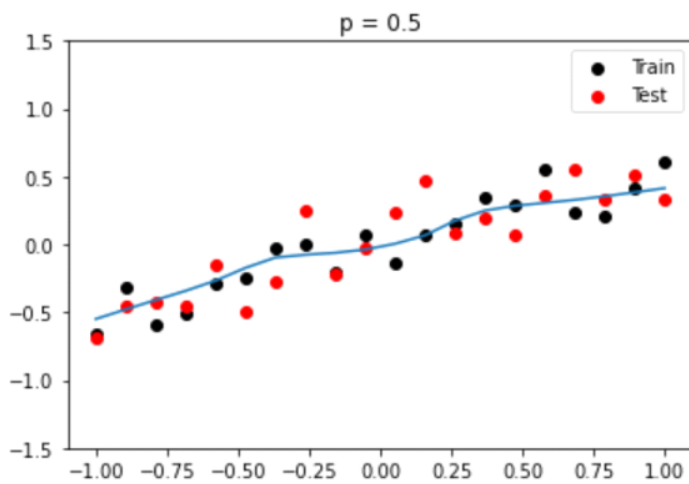
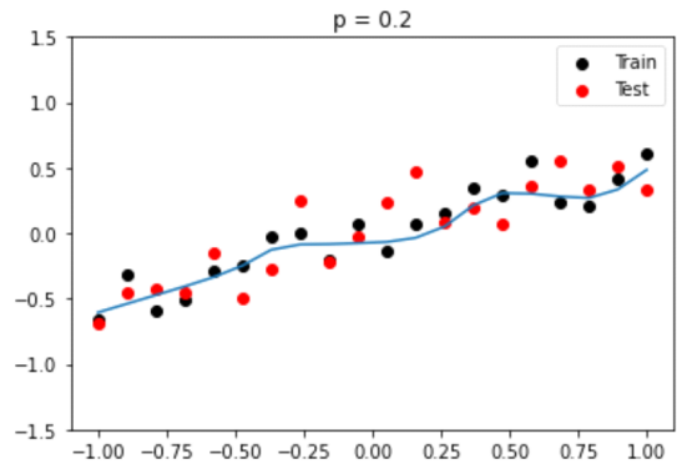
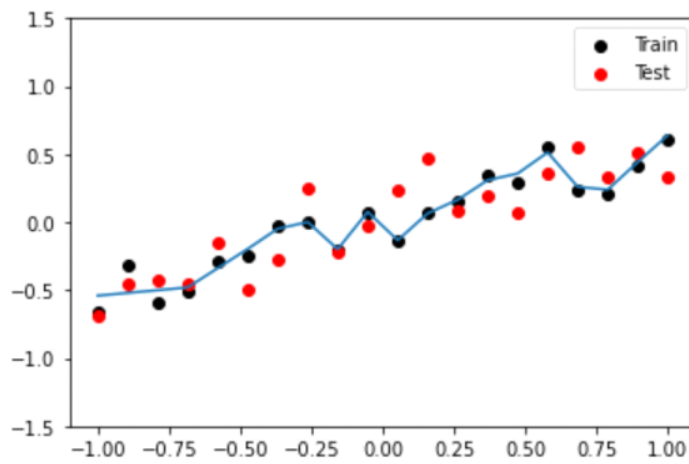
12 May 2022 09:34

Classification Code Example

12 May 2022 09:34

Effect of p

12 May 2022 10:06

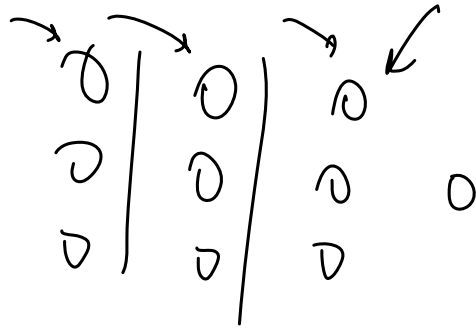


Practical Tips and Tricks

12 May 2022 09:34

1) Overfitting $P \uparrow$, underfitting $P \downarrow$

2) Last layer \rightarrow dropout



3) CNN \rightarrow 40-50-1. (P) \rightarrow ✓
20-30 RNN

ANN \rightarrow 10-50

50

Drawbacks

12 May 2022 09:34

