

ID: 004943505

Name: Calvin Chen

1.

1

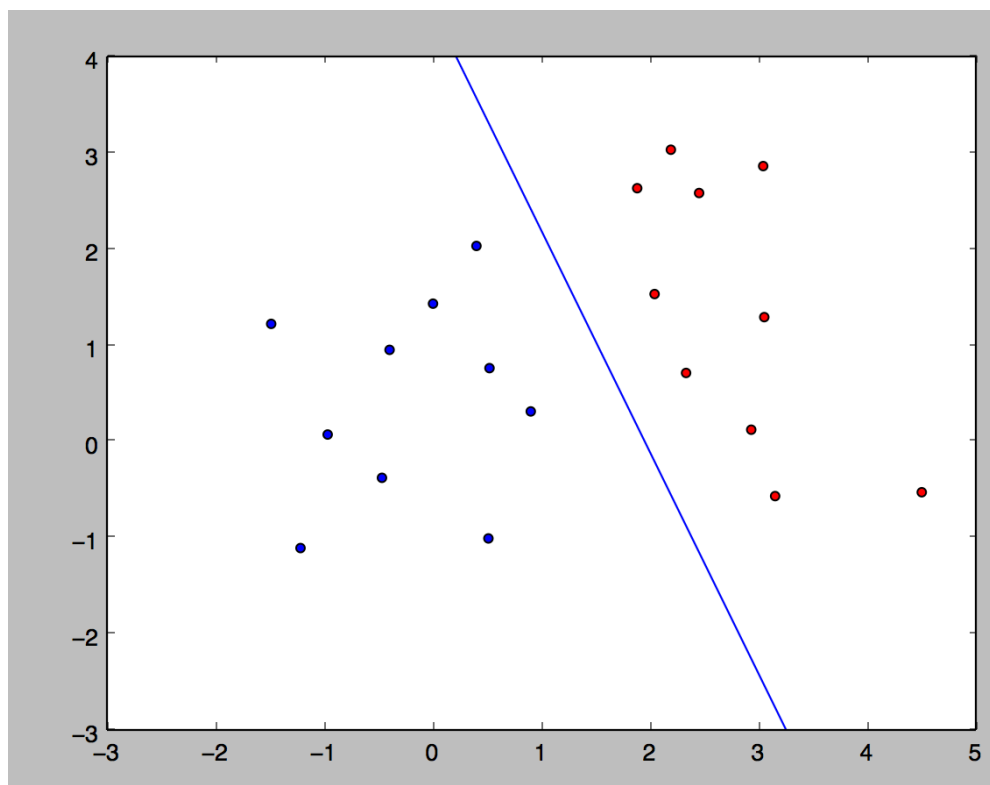
(1) Point 7, Point 18, Point 19

$$\begin{aligned}
 (2) \quad w &= \sum_i \alpha_i y_i x_i = \alpha_7 y_7 x_7 + \alpha_{18} y_{18} x_{18} + \alpha_{19} y_{19} x_{19} \\
 &= 0.9492 \cdot 1 \cdot \begin{pmatrix} 0.91 \\ 0.32 \end{pmatrix} + 0.303 \cdot (-1) \cdot \begin{pmatrix} 2.05 \\ 1.54 \end{pmatrix} + 0.9053 \cdot (-1) \cdot \begin{pmatrix} 2.34 \\ 0.72 \end{pmatrix} \\
 &= \begin{pmatrix} -1.87578 \\ -0.814692 \end{pmatrix}
 \end{aligned}$$

$$\begin{aligned}
 (3) \quad b &= \sum_{k: \alpha_k > 0} (y_k - w^T x_k) / N_k = \frac{1}{3} (1 - (-1.87578 - 0.814692) \begin{pmatrix} 0.91 \\ 0.32 \end{pmatrix}) - 1 - (-1.87578 - 0.814692) \begin{pmatrix} 2.05 \\ 1.54 \end{pmatrix} \\
 &\quad - 1 - (-1.87578 - 0.814692) \begin{pmatrix} 2.34 \\ 0.72 \end{pmatrix} \\
 &= \frac{1}{3} (11.04354) = 3.6812
 \end{aligned}$$

$$(4) \quad f(x) = -1.87578 x_1 - 0.814692 x_2 + 3.6812$$

$$(5) \quad f(x) = -1.87578(-1) - 0.814692(2) + 3.6812 = 1.9296$$



2.

(1)

$$(p+1)3 + (3+1)4 + (4+1)k = 3p + 5k + 19$$

(2)

initialize all weights W_{kj} and biases θ_j to small random number

repeat until terminating condition meets:

for each training data x :

Propagate inputs forward

Backpropagate the error

terminating condition

Net Input and Output Calculations

$$I_j = \sum_k W_{kj} O_k + \theta_j$$

$$O_j = \frac{1}{1 + e^{-I_j}}$$

Unit, j	Net Input, I_j	Output, O_j
3	$0.4 + 0.2 = 0.6$	$1/(1 + e^{-0.6}) = 0.646$
4	$-0.1 + (-0.4) = -0.5$	$1/(1 + e^{-0.5}) = 0.378$
5	$-0.2 \times 0.646 - 0.3 \times 0.378 + 0.1$ $= -0.143$	$1/(1 + e^{-0.143}) = 0.464$

Calculation of the error at each node

$$T = 1 \quad \eta = 0.8$$

Unit, j	Err_j
5	$0.464(1 - 0.464)(1 - 0.464) = 0.133$
4	$0.378(1 - 0.378)(-0.3)(0.133) = -0.009$
3	$0.646(1 - 0.646)(-0.2)(0.133) = -0.006$

$$j=5 \Rightarrow Err_5 = O_5(1 - O_5)(T - O_5)$$

also

$$Err_j = O_j(1 - O_j) \sum_k Err_k W_{jk}$$

$$W_{ij} = W_{ij} + \eta \epsilon_{rj} O_i$$

$$\theta_j = \theta_j + \eta \epsilon_{rj}$$

Calculations for weight and bias updating

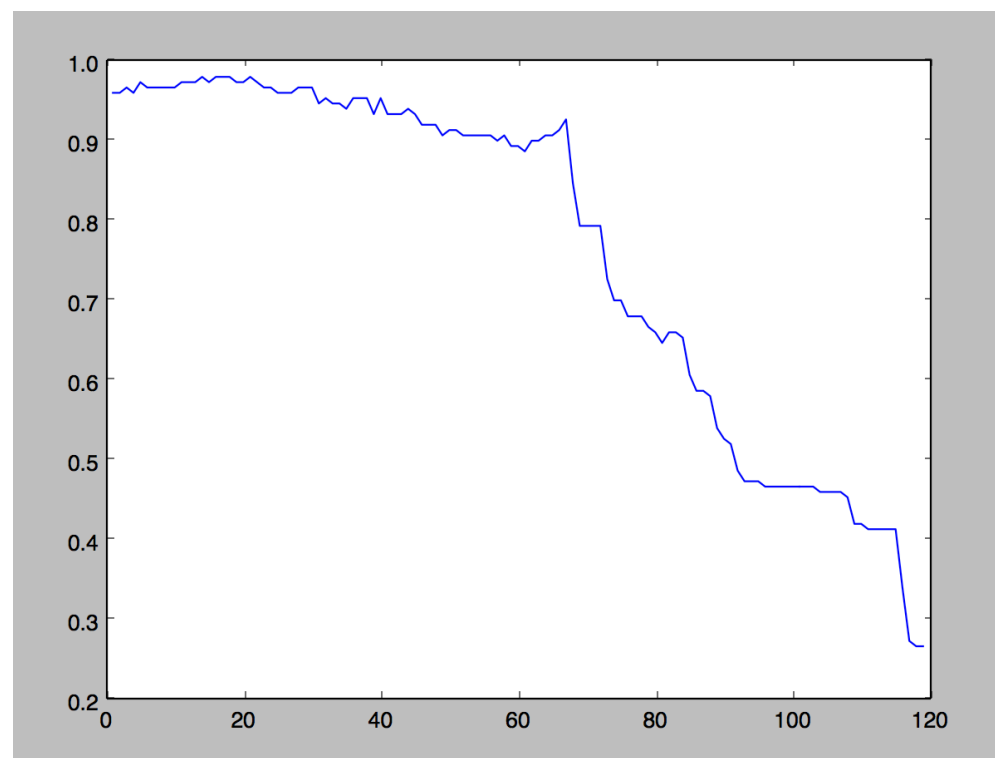
Weight or Bias	New Value
w_{35}	$-0.2 + 0.8 \cdot 0.133 \cdot 0.666 = -0.131$
w_{45}	$-0.3 + 0.8 \cdot 0.133 \cdot 0.378 = -0.260$
w_{13}	$-0.3 + 0.8 \cdot (-0.006) \cdot 0 = -0.3$
w_{14}	$0.2 + 0.8 \cdot (-0.009) \cdot 0 = 0.2$
w_{23}	$0.4 + 0.8 \cdot (-0.006) \cdot 1 = 0.395$
w_{24}	$-0.1 + 0.8 \cdot (-0.009) \cdot 1 = -0.107$
θ_5	$0.1 + 0.8 \cdot 0.133 = 0.206$
θ_4	$-0.4 + 0.8 \cdot (-0.009) = -0.407$
θ_3	$0.2 + 0.8 \cdot (-0.006) = 0.195$

3.

(1) Best value of K = 14

(2)

```
Average accuracies with 5-fold cross validation for K varying from 1 to 119:  
[ 0.96      0.96      0.96666667 0.96      0.97333333 0.96666667  
 0.96666667 0.96666667 0.96666667 0.96666667 0.97333333 0.97333333  
 0.97333333 0.98      0.97333333 0.98      0.98      0.98  
 0.97333333 0.97333333 0.98      0.97333333 0.96666667 0.96666667  
 0.96      0.96      0.96      0.96666667 0.96666667 0.96666667  
 0.94666667 0.95333333 0.94666667 0.94666667 0.94      0.95333333  
 0.95333333 0.95333333 0.93333333 0.95333333 0.93333333 0.93333333  
 0.93333333 0.94      0.93333333 0.92      0.92      0.92  
 0.90666667 0.91333333 0.91333333 0.90666667 0.90666667 0.90666667  
 0.90666667 0.90666667 0.9      0.90666667 0.89333333 0.89333333  
 0.88666667 0.9      0.9      0.90666667 0.90666667 0.91333333  
 0.92666667 0.84666667 0.79333333 0.79333333 0.79333333 0.79333333  
 0.72666667 0.7      0.7      0.68      0.68      0.68  
 0.66666667 0.66      0.64666667 0.66      0.66      0.65333333  
 0.60666667 0.58666667 0.58666667 0.58      0.54      0.52666667  
 0.52      0.48666667 0.47333333 0.47333333 0.47333333 0.46666667  
 0.46666667 0.46666667 0.46666667 0.46666667 0.46666667 0.46666667  
 0.46666667 0.46      0.46      0.46      0.46      0.45333333  
 0.42      0.42      0.41333333 0.41333333 0.41333333 0.41333333  
 0.41333333 0.34      0.27333333 0.26666667 0.26666667]  
Best value of K:  
14
```



(3)

When K = 1, it would be easily affected by noise.

When K = 120, it would always return the label with most data point.

Thus, the K I found is the best value since it can deal with noise any do not serve all the data point.