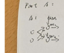
UID: 3036333386 Name: SHI Yaqing.

Part A:

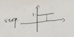
A1 

$(0,0) = 0\times1 + 0\times1 - 1.5 = -1.5 \xrightarrow{step} 0$

$(1,0) = 1\times1 + 0\times1 - 1.5 = -0.5 \rightarrow 0.$

$(0,1) = 0\times1 + 1\times1 - 1.5 = -0.5 \rightarrow 0.$

$(1,1) = 1\times1 + 1\times1 - 1.5 = 0.5 \rightarrow 1.$



A2:

AND: $\begin{bmatrix} w_1 = 1 \\ w_2 = 1 \\ w_0 = -1.5 \end{bmatrix}$. $\Rightarrow \begin{cases} w_1 + w_2 + w_3 > 0 \\ w_2 + w_3 < 0 \\ w_1 + w_3 < 0 \\ w_3 < 0 \end{cases} \Rightarrow w = (1, +1, -1)^T.$

NOT $\begin{bmatrix} w_1 = -2 \\ w_0 = 1 \end{bmatrix} \Rightarrow \begin{cases} w_1 + w_2 < 0 \\ w_2 > 0 \end{cases} \Rightarrow w = (-2, 1)^T$

NAND

A	B	OUT
1	1	0
1	0	1
0	1	1
0	0	1

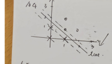
$\Rightarrow \begin{cases} w_1 + w_2 + w_0 < 0 \\ w_1 + w_0 > 1 \\ w_2 + w_0 > 1 \\ w_0 > 1 \end{cases} \Rightarrow w = (-1, -1, 1.5)^T$

NOR

A	B	OUT
1	1	0
0	1	0
1	0	0
0	0	1

$\Rightarrow \begin{cases} w_1 + w_2 + w_0 < 0 \\ w_2 + w_0 < 0 \\ w_1 + w_0 < 0 \\ w_0 > 0 \end{cases} \Rightarrow w = (-1, -1, 0.5)^T.$

A3. The parity problem cannot be learned by a single-layer perceptron. Because a perceptron can only model linearly separable problems. The parity problem is non-linearly separable.

A4 

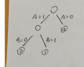
the optimal separating line

support vector is (1,0), (0,1) (1,1), which decide the max margin of the two classes.

A5

Entropy $= -\sum_{i=1}^{n} P(x_i)\cdot(\log_2 P(x_i)) = -0.5\log_2^{0.5} + 4\times(-0.125\times\log_2^{0.125}) = 2$

A6.



decision tree has less computing operation and needs 2 equaling operation.

but perceptron needs 2 times multiplication

A7

1) Split by hair: weighted Gini $= \frac{3}{8}\cdot 0.444 + \frac{3}{8}\cdot 0 + \frac{2}{8}\cdot 0 = 0.16665$ ✓

so we choose split by hair color

2) Split by height: weighted Gini $= \frac{3}{8}\cdot 0.44 + \frac{5}{8}\cdot 0.48 = 0.466$

3) Split by eye: weighted Gini $= \frac{5}{8}\times 0.48 + \frac{3}{8}\cdot 0 = 0.3$

=> If Hair = Dark → Attractive? = No

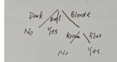
If Hair = Red → Attractive? = Yes

If Hair = Blonde:

If Eyes = Blue → Attractive? = Yes

If Eyes = Brown → Attractive? = No

so the final decision tree is



A8

1). $p = \frac{1}{1+e^{-(w_0 + w_1x_1 + w_2x_2)}} = \frac{1}{1+e^{-(-6+0.05\times40+3.5)}} = \frac{1}{1+e^{0.5}} \approx 0.3775$

2). $p = 0.5 = \frac{1}{1+e^{-z_0}}$ $\therefore z = 0 = w_0 + w_1x_1 + w_2x_2.$

$\cancel{-6x_1 + 3.5 =}$ $-6 + 0.05x_1 + 3.5 = 0$

$x_1 = 50.$

A9.

We choose logistic regression, because the error data of logistic regression is lower than that of 1-nearest neighbors. (36%), and the generalization gap of logistic regression (10%) is much smaller than that of 1-near neighbor (36%), indicating that it can classify new observations more reliably.

A10.

Affected algorithm: KNN, SVM, logistic regression.

Solution: 1) Normalization.

2). Regularization.

A11 ① the number of weak learners.

② higher learning rate.

③ the model complexing of weak learners.

How:

1. decrease the learning rate when the model overfits the dataset.

2. change the loss function.

A12

1. No needs for validation set separated.

2. Computational Efficiency.

3. Stable and Unbiased Error Estimation.

A13 · Hard Voting: Simple and efficient, suitable when probability are unrealiabl-

· Soft Voting: More nuance, when classier provide reliable. probabilities.

· Key Decision Factor: Whether base classifiers support probability outputs.