

HW 3 Report

Part 2.1 Theory

1.

$$f(x) = \begin{cases} 1 & \text{if } \sum_{j=1}^J w_j x_j + b > 0 \\ 0 & \text{if } \sum_{j=1}^J w_j x_j + b \leq 0 \end{cases}$$

The perceptron with a bias term is more expressive than the one without bias because it has an additional parameter, b , that is used in the classifier equation. By accounting for bias, we are able to represent a broader range of classification scenarios, inherently making the classifier more accurate.

2.

- a.
 - i. High classification accuracy
 - ii. High classification accuracy
 - b.
 - i. Low classification accuracy
 - ii. High classification accuracy
 - c.
 - i. High classification accuracy
 - ii. High classification accuracy
 - d.
 - i. Low classification accuracy
 - ii. High classification accuracy
3. The update rule for the bias term b , given learning rate γ and gold label y , when the classifier label doesn't match the gold label during training is $b = b + \gamma * y$. When the classifier label matches the gold label, there is no update rule for b , since the bias stays the same.

Part 3.1 Theory

1.

$$P(c^+ | d) = \frac{P(d | c^+)P(c^+)}{P(d)}$$

2. We need to learn 2^n parameters.

3. We need to learn n parameters.

4. $P(c^+) = P(c^+|d)P(d) / P(d|c^+)$

$$P(c^-) = P(c^-|d)P(d) / P(d|c^-)$$