Assignment8 – Learning

Given: June 15 Due: June 25

Problem 8.1 (Statistical Learning) 0 pt

You observe the values below for 20 games of a sports team. You want to predict the result based on weather and opponent.

Weather	Opponent	Number of wins	Number of losses
Rainy	Weak	3	1
Cloudy	Weak	0	1
Sunny	Weak	4	2
Rainy	Strong	0	2
Cloudy	Strong	2	3
Sunny	Strong	0	2

1. What is the hypothesis space for this situation, seen as an inductive learning problem?
2. Explain whether we can learn the function by building a decision tree.
3. To apply Bayesian learning, we model this situation as a Bayesian network $W \to R \leftarrow O$ using random variables W (weather), O (opponent), and R (game result). What are the resulting entries of the conditional probability table for the cases
   (a) $P(W = rainy) =$ $\boxed{3/10}$
   (b) $P(R = win|O = weak) =$ $\boxed{7/11}$

Solution:

1. The set of functions $\{Rainy, Cloudy, Sunny\} \times \{Weak, Strong\} \to \{Win, Loss\}$.
2. It does not. Even all attributes together, i.e., $Weather$ and $Opponent$, do not determine the result. So no decision tree exists.
3. $P(W = rainy) = 3/10$ and $P(R = win|O = weak) = 7/11$

Problem 8.2 (Neural Networks in Python) 40 pt

Implement neural networks in Python by completing the implementation of `network.py` at `https://kwarc.info/teaching/AI/resources/AI2/network/`.

Hint: You can test your implementation with `test.py`. Note that `test_train_xor_gate` may occasionally fail for a correct solution because it is randomized.

Solution: See `https://kwarc.info/teaching/AI/resources/AI2/network/`.

Problem 8.3 (Support Vectors) 30 pt

Consider the following 2-dimensional dataset

support vector	classification
$\mathbf{x}_1 = \langle 0,0\rangle$	$y_1 = -1$
$\mathbf{x}_2 = \langle 0,0.5\rangle$	$y_2 = -1$
$\mathbf{x}_3 = \langle 0.5,0\rangle$	$y_3 = -1$
$\mathbf{x}_4 = \langle 1,1\rangle$	$y_4 = 1$
$\mathbf{x}_5 = \langle 2,2\rangle$	$y_5 = -1$

1. Give a linear separator in the form $h(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b$ for the dataset without $\mathbf{x}_5$.
2. Explain informally why no linear separator exists for the full dataset of all 5 vectors.
3. Transform the dataset into a 3-dimensional dataset by applying the function $F(\langle u,v\rangle) = \langle u^2, v^2, u+v\rangle$.
4. Give a linear separator for the transformed full dataset in the form $h(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b$.

Solution:

1. Many solutions, e.g., $\mathbf{w} = \langle 1,1\rangle$ and $b = -1$.
2. The points $\mathbf{x}_1, \mathbf{x}_4, \mathbf{x}_5$ lie on a line and the middle one has a different classification than the others. No line can have $\mathbf{x}_1$ and $\mathbf{x}_5$ on one side and $\mathbf{x}_4$ on the other.
3.

support vector $\mathbf{x}$	$F(\mathbf{x})$	classification
$\mathbf{x}_1$	$\langle 0,0,0\rangle$	$y_1 = -1$
$\mathbf{x}_2$	$\langle 0,0.25,0.5\rangle$	$y_2 = -1$
$\mathbf{x}_3$	$\langle 0.25,0,0.5\rangle$	$y_3 = -1$
$\mathbf{x}_4$	$\langle 1,1,2\rangle$	$y_4 = 1$
$\mathbf{x}_5$	$\langle 4,4,4\rangle$	$y_5 = -1$

4. Many solutions, e.g., $\mathbf{w} = \langle -1,-1,2\rangle$ and $b = -1$.

Problem 8.4 (Statistical Learning) 30 pt

We use two observations to determine if it has rained on our property: whether the ground is wet, and whether a bucket we left outside is full.

1. Model this situation as a naive Bayesian network with a boolean class and two boolean attributes.
2. Explain why this network requires 5 parameters ($2n+1$ where $n = 2$ is the number of attributes). Choose 5 names for the parameters and use them to give the conditional probability table of the network.
3. Now assume we have observed for 50 days with the following results:

rain	ground wet	bucket full	number of days
yes	yes	yes	10
yes	yes	no	5
yes	no	yes	6
yes	no	no	4
no	yes	yes	2
no	yes	no	9
no	no	yes	3
no	no	no	11

State the formula for the likelihood of this list of 50 observations in terms of the 5 parameters.

4. Give the Maximum Likelihood approximations for the 5 parameters given these 50 observations. (You just need to compute them, not derive the formula for computing them.)

Solution:

1. We use three boolean variables R (rain), G (ground wet), and B (bucket full) with edges $R \to G$ and $R \to B$.
2. We need
   - one parameter θ for the probability of the class variable R: $P(R = true) = \theta$ and $P(R = false) = 1 - \theta$
   - for each attribute, one parameter for each possible value of the class variable, i.e., 2 if the class variable is boolean:
     - θ_{G1} and θ_{G2} for attribute G: $P(G = true|R = true) = \theta_{G1}$ and $P(G = false|R = true) = 1 - \theta_{G1}$ as well as $P(G = true|R = false) = \theta_{G2}$ and $P(G = false|R = false) = 1 - \theta_{G2}$
     - θ_{B1} and θ_{B2} for attribute B: $P(B = true|R = true) = \theta_{B1}$ and $P(B = false|R = true) = 1 - \theta_{B1}$ as well as $P(B = true|B = false) = \theta_{B2}$ and $P(B = false|R = false) = 1 - \theta_{B2}$
3. Let **d** be the list of observations. The likelihood is

$$P(\mathbf{d}|h_{\theta,\theta_{G1},\theta_{G2},\theta_{B1},\theta_{B2}}) =$$

$$\theta^{10+5+6+4}(1-\theta)^{2+9+3+11}\theta_{G1}^{10+5}(1-\theta_{G1})^{6+4}\theta_{G2}^{2+9}(1-\theta_{G2})^{3+11}\theta_{B1}^{10+6}(1-\theta_{B1})^{5+4}\theta_{B2}^{2+3}(1-\theta_{B2})^{9+11}$$

4. We can derive the formulas by taking logarithms and maximizing. The resulting formulas are the fraction of positive over total examples. So we get

$$\theta = (10 + 5 + 6 + 4)/(10 + 5 + 6 + 4 + 2 + 9 + 3 + 11)$$

$$\theta_{G1} = (10 + 5)/(10 + 5 + 6 + 4) \qquad \theta_{G2} = (2 + 9)/(2 + 9 + 3 + 11)$$

$$\theta_{B1} = (10 + 6)/(10 + 5 + 6 + 4) \qquad \theta_{B2} = (2 + 3)/(2 + 9 + 3 + 11)$$
