Tutorial 12 - Preparatory Questions for the Exam

Question1 [marks]

Solution:

eigenvalues $= \begin{bmatrix} 4 & 2 \end{bmatrix}$, eigenvectors $= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

Question2 [marks]

Solution:

1. We have $P(A) = 4/9$ (4 green bikes out of 9), and $P(A|B) = 3/8$ (since we know that Bob has a green bike, Alice can have one of 3 green bikes out of the remaining 8). Since $P(A) \neq P(A|B)$, the events are not independent.
2. Compute first the probability that Alice and Bob bought bikes of the same color. We have
$P(G,G) = 4/9 * 3/8$, $P(Y,Y) = 3/9 * 2/8$, $P(R,R) = 2/9 * 1/8$
therefore: the probability of buying bikes of different color is:
P(different color) = 1 - P(same color) = $1 - P(G,G) - P(Y,Y) - P(R,R) = 13/18 = 0.722$
3. The requested probability is $P(A \cup B)$. We have
$P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B) - P(A|B)P(B)$
$= (4/9) + (4/9) - (3/8) * (4/9) = 0.722$
4. If Bob bought a green bike, then the conditional probabilities of Alice buying a green, yellow, or red bike are 3/8, 3/8 and 2/8 respectively. The expected amount of money spent by Alice is: $300 * 3/8 + 200 * 3/8 + 100 * 2/8 = 212.50$

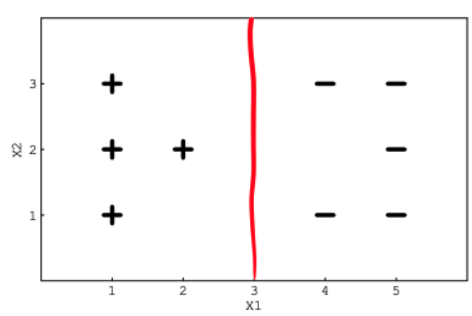
Question3 [marks]

Solution:

- Mean square error.
- $\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{n} \sum_{i=1}^{n} \left(h_{\boldsymbol{\theta}}(x^{(i)}) - y^{(i)}\right) x_j^{(i)}$
$\theta_j = \theta_j - \alpha \frac{\partial J(\theta)}{\partial \theta_j} = \theta_j - \alpha \frac{1}{n} \sum_{i=1}^{n} \left(h_{\boldsymbol{\theta}}(x^{(i)}) - y^{(i)}\right) x_j^{(i)}$

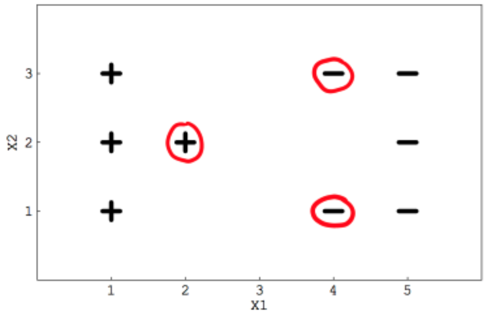
Question4 [marks]

Solution:



1. Because of the large C value, the decision boundary will classify all of the examples correctly. Furthermore, among separators that classify the examples correctly, it will have the largest margin (distance to closest point).

2. $x_1 - 3 = 0$, $w = [1, 0]$, $b = -3$



3. These examples are the support vectors; all of the other examples are such that their corresponding constraints are not tight in the optimization problem, so removing them will not create a solution with smaller objective function value (norm of w). These three examples are positioned such that removing any one of them introduces slack in the constraints, allowing for a solution with a smaller objective function value and with a different third support vector; in this case, because each of these new (replacement) support vectors is not close to the old separator, the decision boundary shifts to make its distance to that example equal to the others.

Question5 [marks]

Solution: We may overfit by directly maximising the likelihood of logistic regression if (1) the training sample is small or (2) the model is too complex. Even in data rich settings, we prefer regularised models (in similar vein to regularising linear regression models with methods such as ridge regression, lasso, elastic net etc). Optimising with respect to MLE may lead to brittle solutions that do not generalise well. For more details, please see Murphy 8.3.6. In logistic regression, we prefer MAP estimation to MLE estimation. We may overfit if we use MLE, placing too much probability mass on our training data and creating solutions that do not generalise well to test data. Murphy explains in his book: Suppose we have linearly separable data. MLE is obtained when $|w| \to \infty$, which is an infinitely steep sigmoid function $w2x > w0$...linear threshold unit. This assigns maximal amount of probability mass to the training data, and accordingly will overfit and not generalise well. To prevent this, we can use l2 regularisation.

Question6 [marks]

Solution:

$$p(X, Z|Y) = \frac{p(X, Y, Z)}{p(Y)} = \frac{p(Z|Y)p(Y|X)p(X)}{p(Y)} = p(Z|Y)\frac{p(Y|X)p(X)}{p(Y)} = p(Z|Y)P(X|Y) \qquad (1)$$

Question7 [marks]

Solution: Base on the distance between unknown point to 3 closest neighbors. (class red)