1 Review Questions

1. Assume the ...

   - What is the ...
     $$(1 - \lambda)^k \lambda$$

   - What is the expected ...
     Let k be the number of tosses needed to get the first head. Then the expected number of k is defined by

     $$
     \begin{align}
     E[k] &= \lambda \times 1 + (1 - \lambda)\lambda \times 2 + (1 - \lambda)^2 \lambda \times 2 + \cdots \tag{1} \\
     (1 - \lambda)E[k] &= (1 - \lambda)\lambda + (1 - \lambda)^2 \lambda \times 2 + (1 - \lambda)^3 \lambda \times 3 + \cdots \tag{2}
     \end{align}
     $$

     By subtracting (2) from (1), we get

     $$
     \begin{align}
     \lambda E[k] &= \lambda + (1 - \lambda)\lambda + (1 - \lambda)^2 \lambda + (1 - \lambda)^3 \lambda + \cdots \\
     &= \lambda \frac{1}{1 - (1 - \lambda)} \\
     &= 1.
     \end{align}
     $$

     Thus, $E[k] = 1/\lambda$.

2. Answer to problem 2

   - What is the partial ...
     $$\frac{\partial f}{\partial x} = 6x - y - 11, \ \ \frac{\partial f}{\partial y} = 2y - x$$

   - Find a point ...
     $$
     \begin{cases}
     6x - y - 11 = 0 \\
     2y - x = 0
     \end{cases}
     $$
     By solving these equations, we get $(x, y) = (2, 1)$.

3. Answer to problem 3

   - Assume that The hyperplane has it normal vector ω, and it is away from the origin by $-b/\|\omega\|$. The example of a hyperplane for $n = 2$ and $n = 3$ is in Figure 1.
   - Assume we have two
     $$\left| \frac{-b_1}{\|\omega\|} - \frac{-b_2}{\|\omega\|} \right| = \frac{|b_1 - b_2|}{\|\omega\|}$$

2 Basic Concepts

1. Define in one sentence: (1) training set, (2) test set, (3) validation set.

   - training set

     Training set is a set of data used to optimize a hypothesis function.

   - test set

     Test set is a set of real-world data used to measure the accuracy of the hypothesis generated through training and validation phases.

   - validation set

     Validation set is a set of data used to estimate the performance of the hypothesis.

2. Can you use the validation set as a test set?

   No. Since validation set is used to estimate the accuracy of the hypothesis during the validation phase, the resulting hypothesis is optimized for the validation set, and it is meaningless to use the validation set as a test set in order to measure the actual performance for the real-world data.

3. Define in one sentence: overfitting

   A hypothesis is said to overfit the training data if it has smaller error on the training data but loses the generalization performance and has larger error on test data.

4. True or False (and why):

3 Decision Trees