

**Please note:** in AML, **Exercises** are embedded in the text of each section; **Problems** are at the end of each chapter. Please be sure you work the assigned problem or exercise to get credit.

1. AML Problem 1.7(a) (p. 36), plus the following part:
  - (b) Take the scenario of part (a), for the case of 1,000 coins, and  $\mu = 0.05$ . Consider the following interpretation in applying it in a machine learning setting.

There is one hypothesis that is given (one decision boundary and corresponding set of decision regions, or one decision rule); call it  $h$ . The out of sample error is  $E_{out}(h) = 0.05$ , and the in-sample error depends on the dataset drawn.

**Hint:** The number of tosses of a coin,  $N = 10$ , corresponds to the size of a dataset.

Complete the machine-learning interpretation by answering the following:

- (i) What do the 1000 coins represent?
- (ii) What does the calculation in part (a), for 1000 coins and  $\mu = 0.05$ , represent?

**Tip:** if you are not sure of your answers, try working Problem 2, then come back to this problem.

2. Suppose you have trained a model in a binary classification problem, and you want to estimate its error based on a test set.

You want to estimate this error empirically for the case where labeled data is expensive. So you decide to first do some simulations to see how your test-set error might vary with the “luck of the draw” of the test set.

Let the true probability of error on your model be  $E_{out}(h) = \mu$ . Because  $\mu$  is unknown to you, for purposes of simulation you will try different values of  $\mu$ .

**Method:** Conceptually, a test set is  $\mathcal{D}^{(N)}$  created by drawing  $N$  data points randomly from the input space, with replacement. An expert could correctly label each data point, and then you could color each data point as “correct” or “incorrect” depending on the ML classifier’s prediction.

You decide to simulate this by drawing (colored) data points randomly, with replacement, from a bin of “correct” and “incorrect” data points, with  $P(\text{incorrect}) = \mu$ .

- (a) Let  $\mu = 0.20$ .
- Draw a colored dataset  $\mathcal{D}^{(10)}$  of size  $N = 10$ . From your 10 drawn data points, compute the error rate  $E_{\mathcal{D}^{(10)}}(h)$ . Is it equal to 0.20? Explain why or why not.
  - Calculate theoretically, using the given value of  $\mu$ , the probability  $P(E_{\mathcal{D}^{(10)}}(h) = \mu)$  for any draw of  $N$  data points, rounded to 2 decimal places (0.xx).
- (b) Repeat the experiment of part (a)(i) 100 times. Keep a record of the error rate  $E_{\mathcal{D}^{(10)}}(h)$  from each run (no need to turn in these 100 values). Give the following statistics and answer these questions:
- $\max\{E_{\mathcal{D}^{(10)}}(h)\},$   
 $\min\{E_{\mathcal{D}^{(10)}}(h)\},$   
sample mean  $\{E_{\mathcal{D}^{(10)}}(h)\},$   
sample standard deviation  $\{E_{\mathcal{D}^{(10)}}(h)\}.$
  - How many of your 100 runs had an error rate different than  $\mu$ ? Does this agree with the value of  $P(E_{\mathcal{D}^{(10)}}(h) = \mu)$  from item (a)(ii) ?
  - From your results of the 100 runs, give an estimate for the probability
$$P(|E_{\mathcal{D}^{(10)}}(h) - \mu| < 0.05)$$
- (c) Repeat parts (a)(ii) and (b) for the following parameters:
- $\mu = 0.20, N = 100$
  - $\mu = 0.50, N = 10$
  - $\mu = 0.50, N = 100$
- (d) Comment on your results:
- How does the accuracy of your error-rate estimate from a test dataset, vary with  $N$  ?
  - For the classifier of (c)(ii),(iii): based on the true error rate, did the classifier learn anything?  
How many test datasets of your draws in (c)(ii) and (c)(iii) gave an error rate indicating that the classifier did learn something, assuming that  $E_{\mathcal{D}^{(10)}}(h) \leq 0.45$  means it learned something?

### 3. AML Problem 2.1 (p. 69).

4. [Based on AML Problem 2.2] In 2D feature space, consider the hypothesis set  $\mathcal{H}$  consisting of positive rectangles: each hypothesis is a rectangle-shaped decision boundary, and has value (label) +1 inside and on the rectangle, and value (label) -1 outside. The sides of the rectangle are parallel to the coordinate axes. Answer the following:

(a) Show that  $m_{\mathcal{H}}(3) = 2^3$

(b) Show that  $m_{\mathcal{H}}(4) = 2^4$

**Hint:** Choose a set of data points that are symmetric in some sense; then count dichotomies using symmetry (e.g., show one dichotomy and state how many other dichotomies can be realized in the same way due to symmetry). This can save you some writing and drawing.

(c) You are given that  $m_{\mathcal{H}}(5) < 2^5$ . Give a simple upper bound for  $m_{\mathcal{H}}(N)$ , valid for any  $N$ , that is a polynomial in  $N$ .

**Hint:** AML Theorem 2.4 will not be a considered “simple” polynomial for this problem.

5. [Based on AML Exercise 2.1, p. 45]

(a) Find the smallest break point  $k$  for the hypothesis set consisting of Positive Rays (defined in Example 2.2).

(b) Find the smallest break point  $k$  for the hypothesis set consisting of Positive Intervals (defined in Example 2.2).