6.891 Fall 1999: Problem Set Grading Guide

We are trying out a cross-grading system where each student will grade another student's problem. An answer key will be provided (see below following this guide). The goal of this is to reduce the grading burden on any one person and to allow you to see possibly alternative solutions to the problem. This can help you to better understand the concepts we are trying to convey to the class. We realize this may be the first time you are asked to do this, and we expect questions, so feel free to let us know if you have any comments. You will assign one of four marks to each part of each problem according to the guidelines below. Thank you for helping us out with this system.

For each part of each problem assign one of the following grades:

- $\surd+$: got correct answer and properly justified solution (i.e., "nailed the problem and answer completely correct").
- $\surd$: showed significant work and insight but may not have completely solved the part, or the answer was slightly off (i.e., "got pretty much everything right except made a small error or left off some justification").
- $\surd-$: showed significant work and insight but had a some gaps in justification or answer (i.e., "had the right idea, but got lost somewhere or had a major error").
- 0 : did not show significant work or insight, or answer was not applicable to the question.

6.891 Fall 1999: Problem Set #1 Solutions

Pre-questions

Problem 1: Probability, Densities and Bayes' Rule

- (A)

Since $p(x)$ is a continuous density, the probability that $x = 0$ is $Pr(0 < x \leq 0) = \int_0^0 p(x)dx = 0$.

- (B)

As in part (A), the probability that $y = 0$ is $Pr(0 < y \leq 0) = \int_0^0 p(y)dy = 0$.

- (C)

Let ω_0 (ω_1) be the state where the X (Y) process generates a data point. We want $Pr(\omega_0|Z = 0)$. Using Bayes' Rule and working with the probability densities to get the total probability:

$$\begin{aligned}
Pr(\omega_0|Z=0) &= \frac{p(Z=0|\omega_0)Pr(\omega_0)}{p(Z=0)} && (1)\\
&= \frac{p(Z=0|\omega_0)Pr(\omega_0)}{p(Z=0|\omega_0)Pr(\omega_0)+p(Z=0|\omega_1)Pr(\omega_1)} && (2)\\
&= \frac{(0.5)p_X(Z=0)}{(0.5)p_X(Z=0)+(0.5)p_Y(Z=0)} && (3)\\
&= \frac{p_X(Z=0)}{p_X(Z=0)+p_Y(Z=0)} && (4)\\
&= \frac{0.3939}{0.3989+0.2420} && (5)\\
&= 0.6224 && (6)
\end{aligned}$$

where:

$$\begin{aligned}
p_X(x) &= \frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}} && (7)\\
p_Y(y) &= \frac{1}{\sqrt{2\pi}}e^{\frac{-(y-1)^2}{2}} && (8)
\end{aligned}$$

Problem 2: Linear Algebra

- (A)

For all the points to lie on the line, a and b must satisfy the following set of simultaneous equations:

$$\begin{aligned}
ax_1 + by_1 &= 1 && (9)\\
&\vdots && (10)\\
ax_N + by_N &= 1 && (11)
\end{aligned}$$

This can be written much more compactly in the matrix form: $PM = \mathbf{1}$ where $\mathbf{1}$ is an $Nx1$ column vector of 1's.

- (B)

An M satisfying the matrix equation in part (A) will not exist unless all the points are collinear (i.e. fall on the same line). In general, three of more points may not be collinear.

- (C)

If M exists, then we can find it by finding the left inverse of P, but since P is in general not a square matrix P^{-1} may not exist, so we need the pseudo-inverse $(P^TP)^{-1}P^T$. Thus $M = (P^TP)^{-1}P^T\mathbf{1}$.

Some Slightly Harder Problems

Problem 3: High Dimensional Weirdness

- (A)

By the symmetry of the problem, we will take the line that can be characterized by a tangent vector with all positive components. Thus, the dot product of the tangent vector with each of the basis vectors must be same (since the angles are all the same and all coefficients are positive). Furthermore, the length of the tangent vector must be one. This imposes that:

$$\begin{aligned} \mathbf{t}\cdot\mathbf{e_1} &= \mathbf{t}\cdot\mathbf{e_2} = \cdots = \mathbf{t}\cdot\mathbf{e_d} && (12)\\ \mathbf{t}\cdot\mathbf{t} &= 1 && (13) \end{aligned}$$

where $\mathbf{t}$ is the tangent vector and the vector set $\{\mathbf{e_1}, \mathbf{e_2}, \ldots, \mathbf{e_d}\}$ is an orthonormal basis for the space which we will take to be the columns of the identity matrix without loss of generality.

This means that all of the coefficients of $\mathbf{t}$ are the same and their squares sum to one. Therefore, all of the coefficients are equal to $\sqrt{\frac{1}{d}}$. Thus, the cosine of the angle between the line an any axis is $\sqrt{\frac{1}{d}}$ and the angle is $\cos^{-1}(\sqrt{\frac{1}{d}})$.

$$\begin{aligned} \alpha(d) &= \cos^{-1}\sqrt{1/d} && (14)\\ \alpha(2) &= 45^\circ && (15)\\ \alpha(3) &\approx 54.7^\circ && (16)\\ \alpha(256) &\approx 86.4^\circ && (17)\\ \alpha(\infty) &= 90^\circ && (18) \end{aligned}$$

Problem 4: Apples and Worms

For this problem we will introduce two random variables: A and T. A represents whether a worm is in the apple and T represents the result of the test (both can take on one of two values: W [for worm] or N [for no worm]).

Thus, the problem statement defines the following:

$$p(A = w) = 0.002 \quad (19)$$
$$p(T = n|A = w) = 0 \quad (20)$$
$$p(T = w|A = n) = 0.03 \quad (21)$$

- (A)

The unconditional probability that an apple will test as having worms $Pr(T = w)$ can be found by summing the conditional probabilities $Pr(T = w||A = w)$ (Pr(T=w—A=n)) multiplied by their prior probabilities $Pr(A = w)$ ($Pr(A = n)$).

$$\begin{aligned} Pr(T = w) &= Pr(T = w|A = w)Pr(A = w) + Pr(T = w|A = n)Pr(A = n) \\ &= 0.002 \times 1 + 0.03 \times 0.998 \\ &= 0.03194 = 3.194\% \end{aligned} \quad (22)$$

- (B)

To find the probability of of an apple having worms given that it has tested for worms $Pr(A = w|T = w)$, we use Bayes' Rule:

$$\begin{aligned} Pr(A = w|T = w) &= \frac{Pr(T = w|A = w)Pr(A = w)}{Pr(T = w)} \\ &= \frac{1.0 \times 0.002}{0.03194} \\ &\approx 0.062617408 = 6.262\% \end{aligned} \quad (23)$$

- (C)

Combining the classifiers, as suggested by Sun Grocery Stores, is only possible if the tests are not completely dependent (i.e. the tests don't return the same results). If all three companies licensed the same technology (very likely if they have the same performance results), combining the classifiers won't help.

- (D)

However, assuming that they are all independent given the status of the apple, we can test an apple by running it through each test and accepting it if *any* of the tests result in a good classification (since these amazing classifiers have a zero false negative rate). Thus, the new false positive rate of the boosted combination $Pr(T = w|A = n) = 0.03^3$.

- (E)

Cranking through the same set of calculations as in part (B), we now end up with $Pr(A = w|T = w) \approx 0.9867 = 98.67\%$.

- (F)

For this question, we have a group of classifiers all with the following properties:

$$Pr(T = n|A = w) = 0.27 \tag{24}$$
$$Pr(T = w|A = n) = 0.27 \tag{25}$$

We will combine them by having each classifier vote and tally the votes. The final classification will be made by comparing the result of the tally to a threshold. Note that this is the best we can do. The vote tally contains all of the information we could get from the individual tests given that they are independent and identical. Let's let n be the number of classifiers and θ be the threshold. If S is the random variable of the number of classifiers voting for "worm," we know that S is distributed as a binomial distribution. This gives that (now letting T be the value for the boosted group classifier by an abuse of notation)

$$\begin{aligned}
Pr(T = n|A = w) &= \sum_{t=0}^{\theta-1} Pr(S = t|A = w) \\
&= \sum_{t=0}^{\theta-1} 0.73^t 0.27^{n-t} \binom{n}{t} \qquad (26) \\
Pr(T = w|A = n) &= \sum_{t=\theta}^{n} Pr(S = t|A = n) \\
&= \sum_{t=\theta}^{n} 0.27^t 0.73^{n-t} \binom{n}{t} \qquad (27)
\end{aligned}$$

Note that if you are unfamiliar with the binomial distribution, the above equations can be derived simply by counting. We wish to know the probability that *at least* θ of the n classifiers voted for a worm. That is the sum of t from θ to n of the chance that exactly t classifiers voted for worm. There are n choose t different votings that will give exactly t worm votes and each has probability $0.27^t 0.73^{n-t}$ (since we are assuming the apple actually has no worm).

The following program solves this problem by brute-force:

```
% prob4F(baserate, targetfp, targetfn
% solves problem #4F on problem set #1 for 6.891, fall 1999
% For that problem, the proper inputs are (.73,.01,.0001)
function prob4(baserate, targetfp, targetfn)

n=1;
while(1),
   for t=0:n,
      falsen = 0;
      for i=0:t-1,
         falsen = falsen + ...
            baserate^i * (1-baserate)^(n-i) * nchoosek(n,i);
      end;
      falsep = 0;
      for i=t:n,
         falsep = falsep + ...
            (1-baserate)^i * baserate^(n-i) * nchoosek(n,i);
      end;
      fprintf(1,'%d / %d:  (%f,%f)\n',n,t,falsen,falsep);
      if (falsen < targetfn & falsep < targetfp),
         return;
      end;
   end;
   n = n+1;
end;
```

Running it with the parameters from the problem yields the results that Sun Groceries needs 41 classifiers with a threshold set at 19.