Advanced topics in Machine Learning
Assignment 1

Mariuta Nicolae (rqt629) September 13, 2015

Question 1

I assign random variable X_i with sample space $X_i = \{0,1\}$ where value 0 means that passenger did not show up and 1 means that the passenger has shown up.

Since 5% of the 10000 passengers did not show up, it means that 9500 have shown up so we can estimate the expected value for 100 passengers:

$$E\left[\sum_{i=1}^{100} X_i\right] = 95$$

By applying the Hoeffding's inequality, I obtain the bound:

$$P\left[\sum_{i=1}^{100} X_i - E\left[\sum_{i=1}^{100} X_i\right] > \ \varepsilon\right] \leq \exp(-\frac{2\varepsilon^2}{\sum_{i=1}^{100}(1-0)^2})$$

Because we want to calculate the bound the probability that the number of people that show up is larger than the number of seats, we have that $\varepsilon = 99 - 95 = 4$

$$P\left[\sum_{i=1}^{100} X_i - 95 > \ 5\right] \leq \exp(-\frac{2 * 4^2}{\sum_{i=1}^{100}(1-0)^2})$$

$$P\left[\sum_{i=1}^{100} X_i \geq 99\right] \leq \exp\left(-\frac{32}{100}\right) = 0.726$$

As a comparison, if I use Markov's inequality to bound the same probability, I obtain:

$$P\left(\sum_{i=1}^{100} X_i \geq 99\right) \leq \frac{E\left[S\sum_{i=1}^{100} X_i\right]}{99} = \frac{95}{99} = 0.9595$$

Question 2

For this problem, the objective function is the Euclidean distance between (x,y) and the point (-2,2):

$$f(x,y) = \sqrt{\left(x-(-2)\right)^2 + (y-2)^2}$$

And after doing some basic calculation, I could bring that to the form:

$$f(x,y) = \sqrt{x^2 + y^2 + 4(x - y)}$$

So $f(x,y)$ that will be the function we want to minimize, but there is the constraint that the point (x,y) to be inside the circle with radius 1 around the origin, so we also have the constraint function :

$$x^2 + y^2 = 1 \quad => x^2 + y^2 - 1 = 0$$

So given this optimization problem, I define the Lagrangian function by combining the objective and the constraint function:

$$F(x,y,\lambda) = \sqrt{x^2 + y^2 + 4(x - y)} + \ \lambda(x^2 + y^2 - 1)$$

Now, I calculate 3 partial derivatives of $F(x,y,\lambda)$ with respect to x, y and λ.

Since it is a quite tricky function to calculate derivative and to avoid doing mistakes, I used an online tool to calculate the derivatives with respect to x and y. I obtained:

$$\frac{\partial F(x,y,\lambda)}{\partial x} = \frac{2x + 4}{2\sqrt{x^2 + 4x + y^2 + 4y}} + 2\lambda x$$

$$\frac{\partial F(x,y,\lambda)}{\partial y} = \frac{2y + 4}{2\sqrt{x^2 + 4x + y^2 + 4y}} + 2\lambda y$$

$$\frac{\partial F(x,y,\lambda)}{\partial \lambda} = x^2 + y^2 - 1$$

For solving the optimization problem, I obtain a system with 3 equations to be solved:

$$\begin{cases} \dfrac{2x + 4}{2\sqrt{x^2 + 4x + y^2 + 4y}} + 2\lambda x \ = 0 \\ \dfrac{2y + 4}{2\sqrt{x^2 + 4x + y^2 + 4y}} + 2\lambda y = 0 \\ x^2 + y^2 - 1 = 0 \end{cases}$$

The optimal solution is found by solving the above system of equation, but I unfortunately did not manage to solve that one.

I am not sure how correct is, but I think I can approach the problem in a simpler way so I can make the calculations easier. Since we want to minimize the objective function:

$$f(x,y) = \sqrt{x^2 + y^2 + 4(x - y)}$$

I am also going to minimize the result of what is below the square root, so I will reduce to the function:

$$f(x,y) = x^2 + y^2 + 4(x-y)$$

that is also going to minimize the Euclidean distance.

By using the same constraint function as previously, the Lagrangian function becomes:

$$F(x,y,\lambda) = x^2 + y^2 + 4(x-y) + \lambda(x^2 + y^2 - 1)$$

And now it is a lot easier to calculate the derivatives:

$$\frac{\partial F(x,y,\lambda)}{\partial x} = 2x + 4 + 2\lambda x$$

$$\frac{\partial F(x,y,\lambda)}{\partial y} = 2y - 4 + 2\lambda y$$

$$\frac{\partial F(x,y,\lambda)}{\partial \lambda} = x^2 + y^2 - 1$$

With that simplification, I also get a system of equations that is a lot easier to solve:

$$\begin{cases} 2x + 4 + 2\lambda x = 0 \ => x = \frac{2}{1+\lambda} \\ 2y - 4 + 2\lambda y \ = 0 => y = -\frac{2}{1+\lambda} \\ x^2 + y^2 - 1 = 0 => \frac{4}{(1+\lambda)^2} + \frac{4}{(1+\lambda)^2} - 1 = 0 \end{cases}$$

Now, I can calculate the value of λ:

$$1 = \frac{8}{(1+\lambda)^2} \quad => \lambda^2 + 2\lambda + 1 = 8 => \lambda^2 + 2\lambda - 7 = 0$$

In the next step, I calculate the possible values of λ:

$$\lambda = 1.82 \ \ or \ \ \lambda = -3.82$$

By making the replacements, I obtain the possible values for λ, x and y I obtain:

$$(\lambda \ = 1.82\,; x = 0.71; y = -0.71) \ \ or(\lambda \ = -3.82\,, x = -0.71; y = 0.71)$$

If I look at the graphic with the circle and the second pair of values for x and y($x = -0.71; y = 0.71$) I see these are exactly the coordinates of the closest point on the circle to our objective (-2,2). The other pair of coordinates that I obtained are symmetric on the other side of y axis.

Question 3

1. Learning with H_{d}

For deriving this high probability bound, because H_{d} is finite, I use the bound for the finite hypothesis classes:

$$P\left\{\exists h \in H_{\mathrm{d}}:\ L(h) > L'(h,S) + \frac{\sqrt{\ln\left(\frac{M}{\delta}\right)}}{2n}\right\} \leq \delta$$

Where h is the hypothesis that takes strings from Σ^{d} and returns {0,1} according to the prediction that is made. $L(h)$ is the expected loss of h and $L'(h,S)$ is the empirical loss of h on the training set S that is provided. M is the size of H_{d} and n is the number of samples from the training set. δ is our confidence budget.

With probability greater than $1-\delta$ we have:

$$L(h) \leq L'(h,S) + \frac{\sqrt{\ln\left(\frac{M}{\delta}\right)}}{2n}$$

So we can say that:

$$L(h) - L'(h,S) \leq \frac{\sqrt{\ln\left(\frac{M}{\delta}\right)}}{2n}$$

Because H_{d} is the set of all functions from Σ^{d} to {0,1} then I guess the size of H_{d} is equal with the number of strings with size d from the set which will make that M to have a large value and it will result in a tight bound.

2. Learning with H

Because in this case, the size of hypothesis space is infinite, we will have the bound:

$$P\left\{\exists h \in H:\ L(h) > L'(h,S) + \frac{\sqrt{\ln\left(\frac{1}{p(h)\delta}\right)}}{2n}\right\} \leq \delta$$

For the high probability bound $1-\delta$ we have:

$$L(h) \leq L'(h,S) + \frac{\sqrt{\ln\left(\frac{1}{p(h)\delta}\right)}}{2n}$$

Where we have that $\sum_{h\epsilon H} p(h) \leq 1$

$$L(h) - L'(h,S) \leq \frac{\sqrt{\ln\left(\frac{1}{p(h)\delta}\right)}}{2n}$$

In this case, the bound is a lot wider because the budget is distributed along all the spaces with strings of different sizes.