Math Foundations of ML, Fall 2018

Homework #9

Due Friday November 16, at the beginning of class

As stated in the syllabus, unauthorized use of previous semester course materials is strictly prohibited in this course.

1. Using you class notes, prepare a 1-2 paragraph summary of what we talked about in class in the last week. I do not want just a bulleted list of topics, I want you to use complete sentences and establish context (Why is what we have learned relevant? How does it connect with other things you have learned here or in other classes?). The more insight you give, the better.

2. In this problem, we will try out the James-Stein estimator and see how it compares to the (unbiased) MLE. For $D = 5, 6, 7, 8, 9, 10$, we will take

$$\boldsymbol{\theta}_0 = \frac{2}{\sqrt{D}} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix},$$

and so $\|\boldsymbol{\theta}_0\|_2 = 2$. (For this problem, it really doesn't matter what $\boldsymbol{\theta}_0$ is, only what its norm is.) Our experiment consists of drawing a single random vector

$$X \sim \mathrm{Normal}(\boldsymbol{\theta}_0, \mathbf{I}),$$

and then computing the squared error for the MLE $\hat{\theta}_{\mathrm{MLE}} = X$, and the two James-Stein estimators in Notes 23 (equations 2 and 3 there). Repeat this experiment many, many times and average the squared errors. Report the results for each value of D above, and compare your results to the expressions in the notes (at least for the MLE and the "regular" JS estimator in equation 2).

3. Three friends, Aaron, Blake, and Colin, meet together every week to play poker. They each buy in for \$100, and play until one of them has it all. Poker is a game of skill, but also a game of luck — the winner each week is modeled as a discrete random variable X with distribution parameterized by θ_a, θ_b, with

$$\mathrm{P}(X = A) = \theta_a, \quad \mathrm{P}(X = B) = \theta_b, \quad \mathrm{P}(X = C) = 1 - \theta_a - \theta_b,$$

where

$$\theta_a, \theta_b \geq 0, \quad \text{and} \quad \theta_a + \theta_b \leq 1. \tag{1}$$

Above, event A corresponds to Aaron winning, B corresponds to Blake winning, and C corresponds to Colin winning.

The parameters θ_a and θ_b are unknown, and we want to infer them after observing the winners each week for many weeks. We have no idea of the relative skill of the players at the beginning of this experiment, so our prior is uniform on the triangular region specified by the constraints in (1):

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_a \\ \theta_b \end{bmatrix}, \quad f_\Theta(\boldsymbol{\theta}) = \begin{cases} 2, & \boldsymbol{\theta} \in \mathcal{S}, \\ 0, & \boldsymbol{\theta} \notin \mathcal{S}, \end{cases} \quad \mathcal{S} = \left\{ \boldsymbol{\vartheta} \in \mathbb{R}^2 \ : \ \vartheta[1], \vartheta[2] \geq 0, \ \ \vartheta[1] + \vartheta[2] \leq 1 \right\}.$$

(You might, at this point, want to sketch the set $\mathcal{S}$ in $\mathbb{R}^2$.)

(a) Show that after N weeks, where we have observed N_a wins for Aaron, N_b wins for Blake, and $N_c = N - N_a - N_b$ wins for Colin, the posterior for Θ is given by the *Dirichlet distribution*

$$f_\Theta(\boldsymbol{\theta} | X_1 = x_1, \ldots, X_N = x_n) \propto \theta_a^{N_a} \theta_b^{N_b} (1 - \theta_a - \theta_b)^{N - N_a - N_b}.$$

(The constant in front of the expression on the right turns out to be

$$\frac{\Gamma(N+3)}{\Gamma(N_a+1)\Gamma(N_b+1)\Gamma(N-N_a-N_b+1)},$$

which is the integral of the expression on the right over the constraint set $\mathcal{S}$.)

(b) Using MATLAB (or Python), plot the posterior density if after a year of play, we are at

$$N_a = 30, \quad N_b = 18, \quad N_c = 4.$$

4. Suppose the random variables (X, Y), $X \in \mathbb{R}^2$, $Y \in \{1, 2\}$, have joint distribution given by

$$\mathrm{P}(Y = 1) = \mathrm{P}(Y = 2) = 1/2, \quad f_X(\boldsymbol{x} | Y = y) = \frac{1}{2\pi\sqrt{\det(\boldsymbol{\Sigma}_k)}} \exp\left(-\frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu}_k)^{\mathrm{T}} \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_k) \right),$$

where

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \boldsymbol{\mu}_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \boldsymbol{\Sigma}_1 = \begin{bmatrix} 3 & -6 \\ -6 & 18 \end{bmatrix}, \quad \boldsymbol{\Sigma}_2 = \begin{bmatrix} 16 & -6 \\ -6 & 8 \end{bmatrix}.$$

Draw the regions $\Gamma_1(h^*)$ and $\Gamma_2(h^*)$ that correspond to the Bayes classifier. (You can feel free to use MATLAB or Python for this.)

5. (a) The file `hw09p5data` contains two arrays: `X1` and `X2`. These are samples from an unknown distribution, where `X1` has been assigned "class 1", and `X2` has been assigned "class 2". Implement the nearest neighbor algorithm, and sketch the decision regions Γ_1 and Γ_2 that it defines.

(b) In actuality, the data in the last part was generate using the model from problem 4. Estimate the generalization error $R(h)$ for both the Bayes classifier (problem 4) and the nearest-neighbor rule (part a), and compare the two. This will require the generation of many Gaussian random vectors with specified covariance matrices. Good thing you learned on HW 8 how to do that.