Math Foundations of ML, Fall 2018

Homework #7

Due Friday October 26, 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. As in Problem 4, Homework 6, let

$$\boldsymbol{H} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} -1 \\ -3 \end{bmatrix},$$

   and let $f : \mathbb{R}^2 \to \mathbb{R}$ be

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{x} - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}.$$

   Recreate (using MATLAB or Python), the contour plot of $f(\boldsymbol{x})$ around its minimizer in $\mathbb{R}^2$. On top of the contour plot, trace out the two steps of the conjugate gradient method starting at $\boldsymbol{x} = \boldsymbol{0}$.

3. Write a MATLAB function `cgsolve.m` that implements the method of conjugate gradients. The function should be called as

```
[x, iter] = cgsolve(H, b, tol, maxiter);
```

   where the inputs and outputs have the same interpretation as in the previous problem. Run your code on the $\boldsymbol{H}$ and $\boldsymbol{b}$ in the file `hw6p6_data.mat` (same data as in the last homework) for a `tol` of 10^{-6}. Report the number of iterations needed for convergence, and for your solution $\hat{\boldsymbol{x}}$ verify that $\boldsymbol{H}\hat{\boldsymbol{x}}$ is within the specified tolerance of $\boldsymbol{b}$. Compare against steepest descent.

4. Suppose that two random variables (X, Y) have joint pdf $f_{X,Y}(x, y)$. Find an expression for the pdf $f_Z(z)$ where $Z = X + Y$. You can start by realizing that

$$F_Z(u|X_1 = \beta) = \mathrm{P}\left(Z \le u|X_1 = \beta\right) = \mathrm{P}\left(X_2 \le u - \beta|X_1 = \beta\right).$$

   You can combine the expressions above by integrating over β, and see that the resulting expression corresponds to an integral of $f_{X,Y}(x, y)$ over a half plane. From this, you can get the pdf for Z by applying the Fundamental Theorem of Calculus. How does your expression simplify if X and Y are independent? (Convolution!)

5. Suppose that X_1 and X_2 are Gaussian random variables with

$$\mathrm{E}[X_1] = \mathrm{E}[X_2] = 0, \quad \mathrm{var}(X_1) = \sigma_1^2, \ \mathrm{var}(X_2) = \sigma_2^2, \quad \mathrm{E}[X_1 X_2] = \gamma.$$

Show that for any $w_1, w_2 \in \mathbb{R}$, $Z = w_1 X_1 + w_2 X_2$ is also a Gaussian random variable. (Hint: use your answer to the previous question.) What is the mean and variance of Z?

6. Let $X \in \mathbb{R}^D$ be a Gaussian random vector, $X \sim \mathrm{Normal}(\boldsymbol{\mu}, \boldsymbol{R})$.

   (a) Using your answer to the previous question, show that for and $\boldsymbol{w} \in \mathbb{R}^D$, $Y = \boldsymbol{w}^{\mathrm{T}} X$ is a scalar Gaussian random variable with mean $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\mu}$ and variance $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{w}$.

   (b) Show that if $R_{i,j} = 0$, then X_i and X_j are independent, that is

$$f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2).$$

7. X be a Gaussian random vector taking values in $\mathbb{R}^N$, let E be a Gaussian random vector taking values in $\mathbb{R}^M$, and let $\boldsymbol{A}$ be a $M \times N$ matrix. We have

$$X \sim \mathrm{Normal}(\mathbf{0}, \boldsymbol{R}_x), \quad E \sim \mathrm{Normal}(\mathbf{0}, \boldsymbol{R}_e), \quad X, E \text{ independent.}$$

We will make observation of the random vector

$$Y = \boldsymbol{A}X + E.$$

   (a) From your work above, it is clear that Y is a Gaussian random vector in $\mathbb{R}^M$ and that $\mathrm{E}[Y] = \mathbf{0}$. Find the covariance matrix for the Gaussian random vector $\begin{bmatrix} X \\ Y \end{bmatrix}$ that takes values in $\mathbb{R}^{N+M}$.

   (b) Suppose we observe $Y = \boldsymbol{y}$. What is the minimum mean-square error estimate of X given $Y = \boldsymbol{y}$?

   (c) Suppose $\boldsymbol{R}_x = \sigma_x^2\mathbf{I}$ and $\boldsymbol{R}_e = \sigma_e^2\mathbf{I}$. In this case, your MMSE estimator should look familiar, and you should see immediately that $\hat{x}_{MMSE}$ is in the row space of $\boldsymbol{A}$. What are the $\hat{\alpha}_n$ is the expression below?

$$\hat{\boldsymbol{x}}_{MMSE} = \sum_{n=1}^{N} \alpha_n \boldsymbol{v}_n, \quad \text{where the } \boldsymbol{v}_n \text{ are the right singular vectors of } \boldsymbol{A}.$$

   (d) Take $\boldsymbol{R}_x$ and $\boldsymbol{R}_e$ as in part (c), and assume that $\boldsymbol{A}$ has full column rank. What is MSE $\mathrm{E}[\|\hat{\boldsymbol{x}}_{MMSE} - \boldsymbol{x}\|_2^2]$ of the MMSE estimate $\hat{\boldsymbol{x}}_{MMSE}$?