MATH 312 SUMMER 2019: ASSIGNMENT 4

BENEDICT MORRISSEY

Due: Part I due Monday 5th August, beginning of class. Part II is extra credit due Friday 9th July 3PM (easiest to email it to me).

Note: Use of computers is necessary for Part II of this assignment. We are going to spend a class next week (Monday or Tuesday) on computational tools/methods – probably primarily using R. Provide any code/script that you write (Any language is acceptable). R is freely available. Note that Matlab is available in some UPenn computer labs, there is also an open source analogue Octave.

Note: Answers must be fully explained (unless otherwise specified), using full sentences for full credit. You are encouraged to work on these questions with your classmates, however please write down all people you discussed problems with on your assignment, and *write up all solutions on your own.* Asking these questions on online discussion boards is not acceptable.

1. Part I

(1) (Question 10 of Assighment 3) We denote by $\kappa(T)$ the condition number[1] of T, for an invertible linear function T between inner product spaces.
  (a) Show that $\kappa(TS) \leq \kappa(T)\kappa(S)$ ($S : V \to W$, $T : W \to U$ invertible linear transformations between inner product spaces V, W, and U).
  (b) Write down a matrix with a condition number of 2.
  (c) Write down a matrix with a condition number of 1.
  (d) (Harder) Describe all 2×2 matrices with a condition number[2] of 1.

(2) L^1 Curve fitting via Gradient Descent.
  (a) Show that the function $\|-\|_1 : \mathbb{R}^n \to \mathbb{R}$, $\|\vec{v}\|_1 := \sum_{i=1}^{n} |(\vec{v})_i|$, (where $(\vec{v})_i$ refers to the i^{th} entry of the vector $\vec{v}$), is a norm[3] on $\mathbb{R}^n$. This is called the L^1-norm.
  (b) Show that if we have a norm $\|-\|$ coming from[4] an inner product $\langle -, - \rangle$ on a real vector space V, then we have the identity
$$\langle \vec{x}, \vec{y} \rangle = \frac{\|\vec{x}+\vec{y}\|^2 - \|\vec{x}-\vec{y}\|^2}{4}.$$
  (c) Using the above or otherwise, show that for $n > 1$ there is no inner product $\langle -, - \rangle_1 : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, with the property that the associated norm[5] is the norm $\|-\|_1$ defined above.
  Note: Because of this the minimization procedure via projection we used for least squares curve fitting will *note* work in this case. We can instead try gradient descent. This comment is relevant to problem 1 of Part II of the assignment.
  (d) As with least squares minimization we can reduce curve fitting to minimizing $f(x) = \|A\vec{x} - \vec{b}\|_1$ for some matrix A, and vector $\vec{b}$. If we try doing gradient descent starting with a vector $\vec{x}_0$, what is $(D_{\vec{x}}f)(\vec{x}_0)$?
  Note: The natural last part of this question is the first question in part 2 of this assignment.

[1] Recall that $\kappa(T) = \|T\|\|T^{-1}\|$, where we use the norm of part (b) of Question 9 of assignment 2.

[2] By the condition number of a matrix we mean the condition number of its associated linear transformation.

[3] In the sense of either question 9 of homework 3, or in the sense of the wikipedia page "Normed Vector Space."

[4] In the sense of question 9 a) of assignment 3

[5] ie. The norm produced as in question 9 a) of assignment 3

(3) Use the Metropolis–Hasting Algorithm to the find a Markov chain on three states, with steady state distribution $\vec{p} = \left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right)$.

Check that the Markov chain you produce does indeed have the wanted steady state distribution.

(4) Positive Semi-Definite Matrices.

(a) Let A be a (real) matrix. Show that $B = AA^T$ is symmetric (ie. $B^T = B$), and $\vec{x}^T B\vec{x} \geq 0$ for all $\vec{x} \in \mathbb{R}^n$ (we call this being positive semidefinite).

(b) Show that if B is a real, symmetric, and positive definite (meaning that $\vec{x}^T B\vec{x} \geq 0$ for all $\vec{x} \in \mathbb{R}^n$, with equality if and only if $\vec{x} = 0$) matrix, then we can define an inner product on $\mathbb{R}^n$ defined by

$$\langle \vec{x}, \vec{y} \rangle := \vec{x}^T B\vec{y}.$$

(c) Show that if B is real, symmetric and positive semidefinite then all eigenvalues are real.

(d) Show that eigenvectors corresponding to different eigenvalues of a symmetric, positive semidefinite matrix B are orthogonal (Hint: Use Question 8 of assignment 2, to calculate the dot product $\vec{x} \cdot A\vec{x}$, in two different ways.)

Linear Analogue of Kernel Trick:

(e) (Harder) Let B be a real, $n \times n$, symmetric, positive semidefinite matrix. Show that (in a modification of Question 9 (e) of assignment 2) there is a linear map $\mathbb{R}^n \xrightarrow{p_B} (\mathbb{R}^n)^*$, where $p_B(\vec{v}) = f_{B,\vec{v}} : \mathbb{R}^n \to \mathbb{R}$, defined by

$$f_{B,\vec{v}}(\vec{w}) = \vec{w}^T B\vec{v}.$$

(f) What is the kernel of the map p_B above?

(g) (Harder) For

$$B = \begin{pmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \end{pmatrix}$$

what is the Moore–Penrose pseudoinverse[6] of the map p_B defined in part (e)?

(5) Perron–Frobenius Theory:

(a) Give an example of a Transition matrix for an irreducible Markov chain of period 3.

(b) Give an example of a Transition matrix for a Markov chain of period 2, which is not irreducible.

(c) (Harder) Suppose that a Markov chain is aperiodic. Furthermore suppose that it is *not* irreducible, but we can partition the states into two sets I and J such that:

- For any states $i_1, i_2 \in I$, a system in the state i_1 has a non-zero probability of reaching the state i_2 at some future time step.
- For any states $j_1, j_2 \in J$, a system in the state j_1 has a non-zero probability of reaching the state j_2 at some future time step.
- If a system is in a state $i \in I$, the state has a non-zero probability of being in a state $j \in J$ at some point in the future.

Show that the Markov chain has a unique steady state (equilibrium) probability distribution.

2. Part II

(1) Use a computer to do gradient descent to approximate an L^1-line of best fit[7] $y = ax + b$ to the following data points:

$$(0, 4), (1.1, 4.3), (1.9, 2.7), (3.3, 8.0), (2.5, 6.9).$$

[6] We need a inner product on $(\mathbb{R}^n)^*$. Using the dot product and Question 9 (e) of assignment 2, we get a linear isomorphism $\mathbb{R}^n \to (\mathbb{R}^n)^*$. Let r be the inverse of this map. We then get an inner product on $(\mathbb{R}^n)^* \times (\mathbb{R}^n)^*$, as the composition $(\mathbb{R}^n)^* \times (\mathbb{R}^n)^* \xrightarrow{r \times r} \mathbb{R}^n \times \mathbb{R}^n \xrightarrow{-\bullet-} \mathbb{R}$, where the second map is the dot product on $\mathbb{R}$.

[7] Ie one that minimizes the L^1-norm of the error.

Also find the least squares minimization solution. Plot both lines, together with the data.

(2) Use gradient descent to do logistic regression on the following set of data points, to distinguish those of type A from those of type B:

Type A Data Points	Type B Data Points
$(1, 2)$	$(3, 2.7)$
$(2, 2)$	$(2, 2.3)$
$(3, 3.1)$	$(4, 2)$
$(1, 4)$	$(3.5, 3.6)$
$(2, 3.5)$	$(1.5, 0.5)$

(3) Use a support vector machine to classify the above data rather than logistic regression. Plot the two different hyperplanes corresponding to the two different classifications.

(4) Use a computer to do Monte Carlo simulation (you may use a uniform distribution) to estimate the value of

$$\int_0^1 sin^3(x)cos^3(x)[cos^{-1}(x)tan^{-1}(x)]^4dx$$

Note: By cos^{-1} we mean the inverse cosine function, sometimes denoted by "arccos."