Linear Algebra and Probability Take Home Exam

F Pait

March 2020

Instructions:

- Questions 1, 2, and 3 are the midterm exam. Questions 4 and 5 are additional ones for the final. You may start working on them at any time. An extra question may be included for the final if I judge it will be profitable for your studies—otherwise this is the final exam as well.
- Open book exams. Consult any written sources. Visit the library! Internet use allowed. Feel free to discuss with colleagues, produce your own answers.
- Solve as many problems as you can and find useful. All questions are here for good reasons.
- First draft of work on these problems should be completed Friday, 13th of March (the week after Spring break). We will have opportunities to discuss the problems in class and during office hours. You can continue working on all problems and improve your answers until the 24th of April (end of final exam week).

1. A function $f : \mathbb{R}^n \to \mathbb{R}$ is strictly convex if for all $x, y \in \mathbb{R}^n$ and all $\lambda \in [0,1]$

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y).$$

   A symmetric matrix $P = A^T + A$ is called positive–definite if all its eigenvalues are positive.

   Show that a quadratic function $f(x) = x^T P x$ is a convex function if and only P is positive–definite.

2. You are given sets of data $y_i \in \mathbb{R}^{n_y}$ and $x_i \in \mathbb{R}^{n_x}, i \in 1, 2, ..., n$, and would like to study the conjecture that they are linearly related, that is, $y = Ax$. Strict equality will rarely hold, (even in the unlikely event that the linear model is actually correct, whatever that means) because of experimental error and measurement uncertainty.

   Devise a criterion to judge whether a given value of A is better or worse than another. It should be based on the error accumulated over the n data samples. Ideally, investigate 2 ideas to define adequate criteria: one probabilistic, another deterministic.

   Using your criterion, obtain if possible a formula for the best value of A. Start with the scalar case, $\mathbb{R}^{n_y} = \mathbb{R}^{n_x} = 1$, then try to move to the multidimensional case.

3. Consider a random variable $x_0 \in \mathbb{R}^n$ and a sequence of random variables related by the recursion

$$x_{k+1} = A x_k.$$

   First compute the expectation and variance of x_1 as a function of the expectation and variance of x_0. Then find the expectation and variance of x_k for a general k. What can you say about the asymptotic behavior expectation and variance of x_k, for $k \to \infty$?

   Hint: consider — you guessed! — the eigenvalues and eigenvectors of A.

   Also consider the scalar case, $n = 1$, which helps understand the vector one.

4. Using the convolution formula for the sum of random variables, determine and plot the probability density function of the average of 2 independent random variables which have identical uniform distributions.

   Now find the density of the average of 2 independent instances of the random variable you obtained previously. Iterate the procedure a few times, and observe the resulting probability density.

   To obtain the density of the resulting random variable, use any method you prefer: calculus, numerical or symbolic integration software, a convenient discrete approximation, Monte–Carlo style simulation by drawing random samples and recording histograms, or other.

5. You wish to build a predictive model in the form

$$y = a_1 f_1(x) + a_2 f_2(x) + a_3 f_3(x)$$

for a certain set of data pairs (x_i, y_i), $i \in \{1, 2, \ldots, n\}$, with both x and y scalars. The goal is use readily available information on x to extrapolate, or predict, what the corresponding value of a difficult to measure quantity y will be.

(a) Propose a method for finding the parameter vector $(\hat{a}_1, \hat{a}_2, \hat{a}_3)$ that best fits the observed data, assuming the functions f_1, f_2, and f_3 are arbitrary but know.

(b) Pick some garden–variety functions f_1, f_2, and f_3, which — as you just happen to "know" from your previous experience with the object under study — are appropriate building blocks for your model.

Powers, exponentials, sines, sigmoids, hyperbolic tangents, Gaussians...

(c) Choose the points x_i and simulate the experiments which generate the data using the expression

Pick n, not too large, not too small.

$$y_i = a_1 f_1(x_i + \Delta x_i) + a_2 f_2(x_i + \Delta x_i) + a_3 f_3(x_i + \Delta x_i) + \Delta y_i.$$

Generate the set of $2n$ values for Δx_i and Δy_i as small independent pseudo–random variables using whatever software you choose to work with. What does "small" mean?

In a practical problem the most difficult parts are, of course, obtaining the data and finding reasonable functions. We're ignoring all difficulties by simply doing a computer simulation.

(d) Test the method you proposed using the data you generated. Use the model you obtained to predict the values of y for a few values of x not in the original data set. Plot, display, and interpret the results.

Bingo! You are a data scientist ;-)