Overview

In this third of the course, we focused on linear learning algorithms to model data. To summarize:

1. Background: The SVD and the best basis (questions selected from Ch. 6- Can you fill in the exercises?)

2. How is the rank computed? (theoretically and computationally)

3. The SVD and the pseudo-inverse: How is it computed? Where did we use the pseudo-inverse?

4. Lines of best fit: How did we get the error measures? Give a short derivation of two of the algorithms (error in $y-$coordinate and the median - median line). Be able to give the derivation of the second error function (orthogonally projecting the data to the line).

5. Finding the best linear function:

   (a) How do you change an affine equation into a linear equation?
   (b) Hebb's Rule (the biological version)
   (c) Hebb's Rule (the version with no feedback on p. 6)
   (d) Will the rule on p. 6 converge? (Exercises on p. 6)
   (e) The failure leads to Widrow-Hoff (p. 7)

6. Derivatives (Appendix A)

   Be sure you can linearize different kinds of functions (like the examples, p. 5-6)

   Be able to write a quadratic as $\mathbf{x}^T A\mathbf{x} + \mathbf{b}^T\mathbf{x} + \mathbf{c}$, and take the first and second "derivatives" (e.g., find the gradient and Hessian)- Like exercises 7-8 of the homework.

   Be able to explain the method of gradient descent, and explain by approximately how much we drop (in terms of the function) at each step. You might show it in one dimension.

7. Back to linear learning algorithms: How is it that Widrow-Hoff is (approximately) gradient descent? (Be explicit, starting with the error function).

8. What is translationally invariant data?

9. Best subspaces as feature extraction: If we have p data "points" (really vectors) in $\mathbb{R}^n$, then looking for a small set of "template vectors" (or feature vectors) so that each point is a linear combination of the features is the same as finding the best basis for the data set.

Review Questions

1. We had three lines of best fit- Two of them were designed to minimize error functions- What were the error functions (also show them graphically)?

2. Illustrate the median-median line (you may use a calculator) given the data below:

x	-1	2	1	0	6	3	5	-2
y	5	-1	1	2	-8	-3	-7	7

3. Recall that if we have a matrix B so that $AB = I$ and $BA = I$, then matrix B is called the inverse of matrix A.

   Does the pseudo-inverse of the matrix A, $A^{\dagger}$, satisfy the same properties? Explain (using the SVD):

4. What is Hebb's rule (the biological version)?

5. In pattern classification, suppose I have data in the plane that I want to divide into 5 classes. Would I want to build a pattern classification function f so that the range is the following set:
$$\{1, 2, 3, 4, 5\}$$
   Why or why not? If not, what might be a better range?

6. Given the function $f(x, y)$, show that the direction in which f decreases the fastest from a point (a, b) is given by the negative gradient (evaluated at (a, b)).

7. Illustrate the technique of gradient descent using
$$f(x, y) = x^2 + y^2 - 3xy + 2$$
   (a) Find the minimum.

   (b) Use the initial point $(1, 0)$ and $\alpha = 0.1$ to perform two steps of gradient descent (use your calculator).

8. If
$$f(t) = \begin{bmatrix} 3t - 1 \\ t^2 \end{bmatrix}$$
   find the tangent line to f at $t = 1$.

9. If $f(x, y) = x^2 + y^2 - 3xy + 2$, find the linearization of f at $(1, 0)$.

10. Given just one data point:
$$X = \begin{bmatrix} 2 \\ -1 \end{bmatrix} \qquad T = [1]$$

Initializing W and $\mathbf{b}$ as an appropriately sized arrays of ones, perform three iterations of Widrow-Hoff using $\alpha = 0.1$ (by hand, you may use a calculator). You should verify that the the weights and biases are getting better.

11. If a time series is given by:

$$x = \{1, 2, 0, 3, 4, 5, 2, 1, 0, 3, 4\}$$

Give the result of performing lag 2:

12. If the time series is periodic with period k, what happens when we perform a lag $k-1$?

$$x = \{x_1, x_2, \ldots, x_k, x_1, x_2, \ldots, x_k, \ldots\}$$

13. Be sure you can provide justifications for statements 3-5, p. 96 of Chapter 6 (best basis)- You actually did this for a specific 2-dimensional case in Exam 1.

14. If I know the vector $\mathbf{v}_1$ and the singular value σ_1 from the SVD of a matrix A, can I compute $\mathbf{u}_1$ directly? Was σ_1 really needed?

Matlab Review

Be sure you understand how to write/use a function (with multiple inputs and multiple outputs), and how to write and publish a script file.
Know how to compute and manipulate the vectors coming from the SVD.
Also be sure to complete the lab from Thursday (Nov 5).