Special Topics in Modeling, Exam 1

The in-class exam will be approximately 50 minutes. You may use a calculator to do arithmetic, but for no other purpose (i.e., notes or to run programs), as the exam will be closed book and closed notes.

I will not ask you to memorize the formulas for the lines of best fit (in particular the second one), but you should know how to construct lines using the first and third methods.

Computer arithmetic and error

- Be able to convert numbers back and forth between base 10 and bases 2 and 16 (both integer and fractional parts).
- Know how the IEEE format (just the double format), and its relation to the floating point (normalized) form.
- Know the definition and value of machine ϵ.
- Know which part of the IEEE rounding rule is "special" (not the usual way to round numbers). Why is that there?
- Be able to compute the error using floating point and IEEE rounding rule.

Models

We've been developing several numerical models- The $n-$armed bandit uses a stochastic model (using random numbers), and all of the others are using linear algebra:

- The $n-$armed bandit with several strategies.
- The line of best fit.
- Linear models: A general fit, $y = f(x, a, b, c, \cdots)$ where f is linear in the parameters. We looked at setting up the matrix for $y = ax_1 + bx_1^2 + c\sin(x_2)$, which is linear in a, b, c (but not in x_1, x_2).
- The linear neural network. In this case, we have p pairs of "points", $(\mathbf{x}_i, \mathbf{t}_i)$, where $\mathbf{x}_i \in \mathbb{R}^n$ and $\mathbf{t}_i \in \mathbb{R}^m$, and we try to find matrix W and vector $\mathbf{b}$ so that

  $$W\mathbf{x}_i + \mathbf{b} = \mathbf{t}_i$$

  for each $i = 1, 2, \ldots, p$. To find W and $\mathbf{b}$, we have:

  1. Batch training: Use the least-square solution (In Matlab, the slash command).
  2. Online training: Use the Widrow-Hoff algorithm.

- Linear classification problems: Much the same as the linear model, but in this case, we create the targets by taking the classification into account. In the text, we see that it is better to map to a vector rather than scalars. Once the desired targets have been set up, we can use either batch or online training to find the weights and biases.

- Linear novelty detection: Create both the domain and range vectors to train a linear map to the data (How is this done?). Once the desired domain, range pairs have been created, we can use either batch or online training to find the weights and biases.

The Derivative and Gradient Descent

See the Appendix.

Review Questions

1. Convert the following number from base 10 to base 16: 102.34
2. Convert the following number from base 10 to base 5: 102.34.
3. How big is machine ϵ, and how is it defined (for doubles)?
4. How would you round the following number using the IEEE rounding rule?
$$1.001001001\cdots \times 2^4$$
5. Convert the following base 10 number to binary and express it as a floating point number using the IEEE rounding rule: 44/7
6. Do the following sum by hand using IEEE rounding rule (and double precision):
   (a) $(1 + (2^{-51} + 2^{-52} + 2^{-59})) - 1$
7. What is a mathematical model?
8. Was the $n-$armed bandit problem an example of *supervised* or *unsupervised* learning?
9. What is the greedy algorithm? How was it modified to get the $\epsilon-$greedy algorithm? In particular, how did our average reward depend on ϵ in the test trials (or in the figure that you produced)?
10. What is the "softmax" action selection? In particular, how did we change a set of payouts Q_i to a set of probabilities, P_i?
11. Suppose $Q = [-0.5, 0, 0.5, 1.0]$. Use the softmax selection technique with $\tau = 0.1$ to compute the probabilities.
12. If $Q_1 < Q_2 < Q_3 < Q_4$ for 4 machines, how do the probabilities change (under softmax) as $\tau \to 0$? As $\tau \to 1$?
13. What is the win-stay, lose-shift (or pursuit) strategy? What are the update rules?
14. Suppose we play with three machines, and machine 3 is chosen and gives a big payout (enough to make $Q_t(3)$ the maximum). Update the probabilities for win-stay, lose-shift, if they are: $P_1 = 0.3, P_2 = 0.5, P_3 = 0.2$ and $\beta = 0.3$.

15. In the sample script on page 13, how are the initial probabilities set? (NOTE: I may give you a code "snippet" to interpret, like this question.).

16. Find bases for the column, row and null space for the matrix A below. Also, find the null space of A^T.
$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix}$$

17. Show that, for the line of best fit, the normal equations produce the same equations as minimizing an appropriate error function:

18. Given data:

x	-1	0	1
y	2	1	1

    (a) Give the matrix equation for the *line of best fit.*
    (b) Compute the normal equations.
    (c) Solve the normal equations for the slope and intercept.

19. Use the data in Exercise (18) to find the parabola of best fit: $y = ax^2 + bx + c$. (NOTE: Will you only get a least squares solution, or an actual solution to the appropriate matrix equation?)

20. Use the data in Exercise (18) to find the equation of the median-median line.

21. What is Hebb's rule (the biological version- you can paraphrase)?

22. What is the Widrow-Hoff learning rule? How is it related to Hebb's rule?

23. Let $W = \begin{bmatrix} 1 & 0 & 1 \\ -1 & 1 & 0 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. If $\mathbf{x} = [-1, 0, 1]^T$ and $\mathbf{t} = [2, 3]^T$, use Widrow-Hoff to update $W, \mathbf{b}$ one time using a learning rate of 1 (This is too big of a learning rate to actually use, but it will make your computations easier).

24. Let $\mathbf{x} = [1, 2]^T$. Find the matrix $\mathbf{x}\mathbf{x}^T$, its eigenvalues, and eigenvectors.

25. Show that, if $A = \mathbf{x}\mathbf{x}^T$ for a non-zero vector $\mathbf{x}$, then A has one eigenvalue that is $\|\mathbf{x}\|^2$ and the eigenvector is $\mathbf{x}$. Show that all other eigenvalues are zero by finding the null space of A (think about it in terms of words).

26. Show that the affine mapping: $f(\mathbf{x}) = W\mathbf{x} + \mathbf{b}$ can be written as a linear mapping $\hat{W}\hat{\mathbf{x}}$ for an appropriate $\hat{W}$ and $\hat{\mathbf{x}}$

27. What does "training" mean in terms of our mathematical model?

28. If we use all the data we have at once, what kind of training are we doing? If we learn one data point at a time, what kind of training are we doing?

29. Suppose I have some data in $\mathbb{R}^3$ that belongs to 4 different classes. Do I want my targets to be the real numbers $1, 2, 3, 4$, or are there better ways to build the target values? (Hint: Using $1, 2, 3, 4$ implies that class 1 and class 4 are very far apart- more so than 3 and 4. But this is probably not reflected in the data- Try targets that are equally spaced).

30. Given the function $z = f(x, y)$, 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)).

31. 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).

32. If

$$f(t) = \begin{bmatrix} 3t - 1 \\ t^2 \end{bmatrix}$$

find the tangent line to f at $t = 1$.

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

34. 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.

35. 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, and specify the domain-range pairing we would use in the novelty detection algorithm (same lag).

36. Matlab Questions:

   (a) What's the difference between a script file and a function?

   (b) What does the following code fragment produce?

```
Q=[1 3 2 1 3];
idx=find(Q==max(Q));
```

   (c) What will P be:

```
x=[0.3, 0.1, 0.2, 0.4];
P=cumsum(x);
```