CompSci 527 Midterm Exam Sample— Sample Solution

IMPORTANT: The solutions given below are often much more verbose than what is expected for the exam. The reason for the longer answers is in some cases to help you along should you get stuck, or in other cases to show alternative answers. In the exam, please be telegraphic but clear in your answers. Provide explanations only if they are asked for, or if you feel unsure about your answer, for partial credit.

1. The image C on the right below was obtained by convolving the image I on the left with a 2×2 kernel H. The `'valid'` option was used in Matlab.

   Answer:

   $I =$

1	0	5	0	2	0
0	1	0	0	0	9
0	7	0	3	0	0
1	0	0	0	0	1
2	0	2	0	6	0

   $H =$

2	3
1	1

   $C =$

3	8	5	2	20
15	22	6	9	9
10	7	3	3	2
7	4	6	12	19

   Fill in the four values of the kernel. You may want to briefly explain your reasoning if you are not sure about your answer. [Hint: if you are doing a lot of of calculations, think again.]

   Answer: *The pixel $I(3,4)$ in the input image has value 3, and is surrounded by zeros. Because of this, that pixel is the only nonzero contributor to the values in the 2×2 box*

   $$C(2:3, 3:4) = \begin{bmatrix} 6 & 9 \\ 3 & 3 \end{bmatrix}$$

   of the output image. These four values must be equal to

   $$C(2:3, 3:4) = I(3,4)H = 3H \; ,$$

   that is, three times the kernel. This yields

   $$H = \frac{1}{3} C(2:3, 3:4) = \begin{bmatrix} 2 & 3 \\ 1 & 1 \end{bmatrix} \; ,$$

   as shown above.

   [Note: The trick here is to realize that to solve this "inverse" problem (given input and output, find the kernel) you do not flip the kernel, as you would do for a convolution. If you did not get this answer right, work through the example above: Suppose that H is given, and ask the question "what pixels in C does pixel $I(3,4)$ contribute to, given the formula for convolution?" This may take a bit of patience and a drawing or two to work out, unless you want to reason algebraically (not my preferred option, but it may work for you).]

2. Is the following convolution kernel separable? If so, separate it. If not, prove that it is not.

   $H =$

2	3
1	1

   Answer: *No it is not. If it were, the matrix H would have rank 1 and therefore a zero determinant. Instead,*

   $$\det(H) = 2 \times 1 - 3 \times 1 = -1 \; .$$

   [Note: Different proofs are possible.]

3. What is the gradient of the following function at $x = y = 0$?

   $$f(x,y) = (x-2)^3 \sin y$$

   Answer:

   $$\left[\begin{array}{c} 3(x-2)^2 \sin y \\ (x-2)^3 \cos y \end{array} \right]_{x=y=0} = \left[\begin{array}{c} 0 \\ -8 \end{array} \right]$$

4. Give bases for the row space, null space, range, and left null space of the matrix

$$A = \begin{bmatrix} 3 & 0 & 1 \\ 0 & 1 & 0 \\ 3 & 1 & 1 \\ 3 & -1 & 1 \end{bmatrix}$$

State which basis is for which space.

Answer: *Of course, you do not need to give an explanation in your answer, just the answer itself, if you are sure about it. On the other hand, here is an explanation in case you got stuck.*

All you need to notice is that the third row is the sum of the first two, and the fourth row is the difference of the first two—a realization that is made simple by the various zeros in strategic positions. Since the first two rows are linearly independent, it then follows that the rank of the matrix is two. Therefore the row space has dimension two, and the first two rows (or any two rows) can be used as a basis. For the null space, find any vector that is orthogonal to the first two rows. To make this vector orthogonal to the second row, it must have a zero in second position. To make it orthogonal to the first row, just switch 3 and 1 and change the sign of either. So

$$\textit{row space}(A) = \textit{span}\{\begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\} \quad \textit{and} \quad \textit{null}(A) = \textit{span}\{\begin{bmatrix} 1 \\ 0 \\ -3 \end{bmatrix}\} .$$

Similar reasoning applies to columns. A little trial and error yields a vector $[-2, 0, 1, 1]^T$ orthogonal to the last two columns as a basis for the left null of A. If you cannot come up with that vector by inspection, note that you want some vector $\mathbf{v}$ orthogonal to the second and third column of A, so

$$\begin{bmatrix} 0 & 1 & 1 & -1 \\ 1 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} = 0$$

or

$$v_2 + v_3 - v_4 = v_1 + v_3 + v_4 = 0$$

from which

$$v_2 - v_4 = v_1 + v_4 = 0$$

Pick, say $v_3 = 1$ and $v_4 = 1$, then we need $v_2 = 0$ and $v_1 = -2$ to satisfy the equations above, and this yields the same vector for the left null space of A. Another vector can be found by switching the first two components and changing the sign of one, and doing the same to the last two components, to preserve orthogonality with both the first vector and the basis of the range.

In summary,

$$\textit{range}(A) = \textit{span}\{\begin{bmatrix} 0 \\ 1 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}\} \quad \textit{and} \quad \textit{left null}(A) = \textit{span}\{\begin{bmatrix} -2 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ -1 \\ 1 \end{bmatrix}\} .$$

Answers can vary depending on which rows or columns you choose.

5. Assume that the columns of the matrix

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

are linearly independent. Find expressions for the unit vectors $\mathbf{q}_1$ and $\mathbf{q}_2$ obtained by applying Gram-Schmidt to the columns of this matrix (first column first). Your expressions should contain no variables other than a, b, c, d.

Answer:

$$\mathbf{q}_1 = \frac{1}{\sqrt{a^2 + c^2}} \begin{bmatrix} a \\ c \end{bmatrix} \quad \textit{and} \quad \mathbf{q}_2 = \frac{1}{\sqrt{a^2 + c^2}} \begin{bmatrix} -c \\ a \end{bmatrix} .$$

6. Let (x, y) be a point with real coordinates such that

$$a \leq x < a + 1 \quad \text{and} \quad b \leq y < b + 1$$

for integers a and b. Also let

$$\Delta x = x - a \quad \text{and} \quad \Delta y = y - b .$$

Bilinear interpolation yields an image value at (x, y) that has the following format:

$$\begin{aligned} I(x,y) &= I(a,b)\,(1-\Delta x)\,(1-\Delta y) \\ &+ I(a+1,b)\,\Delta x\,(1-\Delta y) \\ &+ I(a,b+1)\,(1-\Delta x)\,\Delta y \\ &+ I(a+1,b+1)\,\Delta x\,\Delta y \end{aligned}$$

Fill in the missing parts above.

7. On how many bins do Dalal and Triggs settle for orientation histograms in their HoG descriptor?

   Answer: *Nine.*

8. What are "hard examples" in Dalal and Triggs's training methodology for HoG descriptors?

   Answer: *Negative training samples that are misclassified as positive in a first pass of training.*

9. What is a categorical set?

   Answer: *A set that is finite and unordered. "Unordered" here means that the ordering of the elements of the set has no meaning of interest.*

10. Define the misclassification loss function, either in words or by a formula.

    Answer:

$$\mathcal{L}(y, y') = I(y \neq y') = \begin{cases} 1 & \text{if } y \neq y' \\ 0 & \text{if } y = y' \end{cases}$$

11. What is a learner said to do when it outputs a classifier that is 100% accurate on the training data but only 50% accurate on test data, when in fact it could have output one that is 75% accurate on both?

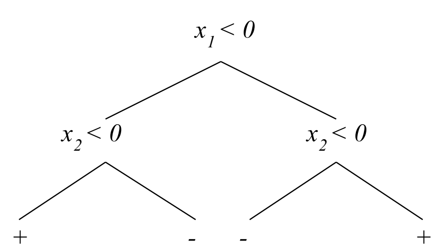
    Answer: *To overfit.*

12. Briefly describe an outcome in dart-throwing that has low variance and high bias.

    Answer: *All darts fall very close to each other (low variance) but far from the target (high bias).*

13. In a certain binary classification problem with two-dimensional features $\mathbf{x} = (x_1, x_2)^T$, all features that have a positive $x_1 x_2$ product are to be classified as positive, and all features that have negative $x_1 x_2$ product are to be classified as negative. Draw an optimal classification tree for this problem that uses individual variables (*i.e.*, either x_1 or x_2) as split variables. Internal nodes in your drawing of a tree have a split criterion, and leaf nodes have a label (either '$+$' or '$-$').

    Answer:



    Your drawing may vary in which node is placed where.

14. In bagging, random sets are drawn uniformly at random out of an initial set T with N elements. How many elements does each random set have?

    Answer: N

15. What information is stored at the leaves of a Hough forest? Define your terminology.

    Answer: *An estimate of the label probability distribution and a list of displacements, one displacement per training sample. A displacement is a vector between the center of a training patch and the center of the object window in the same training image.*

16. Give one possible singular value decomposition for the 2×2 identity matrix I.

Answer:

$$I = U\Sigma V^T = III^T \; .$$

This looks (and is) trivial, but please check that all properties of the SVD are satisfied: $U = I$ is orthogonal, $V = I$ is orthogonal, and $\Sigma = I$ is diagonal with nonnegative elements in non-increasing order on the diagonal. The decomposition is not unique because the two singular values are equal to each other. All SVDs of I are of the form

$$I = UIU^T$$

where U is an arbitrary 2×2 orthogonal matrix.

17. What is data augmentation, and what purpose does it serve?

Answer: *Data augmentation is a technique used by Krizhevsky* et al. *in their 2012 paper to generate several geometric and photometric transformations of each training image when training a deep convolutional net. Each additional image has the same label as the original image it comes from. The new images are added to the training set. The purpose of data augmentation is to increase the size of the training set, thereby improving the generalization error.*

18. A trivial convolutional neural net takes a single, scalar input x, has a first layer with kernel k_1, bias b_1, and nonlinearity $h(a) = a^2$, and a second layer with kernel k_2, bias b_2, and no non-linearity. All kernels and biases are scalar, so the output y is a scalar as well. The loss function is

$$\mathcal{L}(y_n, y) = (y_n - y)^2 \; .$$

Compute (analytically) each of the four components of the gradient of the error $E_1(\mathbf{w})$ for training sample (x_1, y_1) relative to the vector

$$\mathbf{w} = \begin{bmatrix} k_1 \\ b_1 \\ k_2 \\ b_2 \end{bmatrix}$$

of the net's parameters. You may use back-propagation or whatever other means, and introduce intermediate variables if that helps organize the computation. For simplicity, you may want to use the abbreviation

$$e_z = \frac{\partial E_1}{\partial z}$$

for any z. It is OK to give a series of formulas as your answer, as long as every formula uses only variables that were defined before that formula.

Answer: *We could compute derivatives brute-force from the following expression for the error:*

$$E_1(\mathbf{w}) = (y_1 - k_2(k_1 x_1 + b_1)^2 - b_2)^2 \; .$$

to obtain

$$\begin{aligned} e_{k_1} &= -4(y_1 - k_2(k_1x_1 + b_1)^2 - b_2)k_2(k_1x_1 + b_1)x_1 \\ e_{b_1} &= -4(y_1 - k_2(k_1x_1 + b_1)^2 - b_2)k_2(k_1x_1 + b_1) \\ e_{k_2} &= -2(y_1 - k_2(k_1x_1 + b_1)^2 - b_2)(k_1x_1 + b_1)^2 \\ e_{b_2} &= -2(y_1 - k_2(k_1x_1 + b_1)^2 - b_2) \; . \end{aligned}$$

This is an acceptable answer and gets full credit if correct, but the derivation is error-prone and the result is messy.

Alternatively, following the general scheme of back-propagation helps organize the computation. The forward computation is as follows:

$$\begin{aligned} a^{(1)} &= k_1 x_1 + b_1 \\ y^{(1)} &= (a^{(1)})^2 \\ y &= k_2 y^{(1)} + b_2 \\ E_1 &= (y_1 - y)^2 \end{aligned}$$

and we compute derivatives in reverse using the chain rule for differentiation:

$$
\begin{aligned}
e_y &= 2(y - y_1) \\
e_{k_2} &= e_y y^{(1)} \\
e_{b_2} &= e_y \\
e_{y^{(1)}} &= e_y k_2 \\
e_{a^{(1)}} &= 2 e_{y^{(1)}} a^{(1)} \\
e_{k_1} &= e_{a^{(1)}} x_1 \\
e_{b_1} &= e_{a^{(1)}} \ .
\end{aligned}
$$

Just as a sanity check (not required for a good answer), we verify that the two answers above are the same by eliminating all intermediate variables in the last set of expressions:

$$
\begin{aligned}
e_y &= 2(k_2(k_1x_1 + b_1)^2 + b_2 - y_1) \\
\checkmark\ e_{k_2} &= 2(k_2(k_1x_1 + b_1)^2 + b_2 - y_1)(k_1x_1 + b_1)^2 \\
\checkmark\ e_{b_2} &= 2(k_2(k_1x_1 + b_1)^2 + b_2 - y_1) \\
e_{y^{(1)}} &= (k_2(k_1x_1 + b_1)^2 + b_2 - y_1)k_2 \\
e_{a^{(1)}} &= 4(k_2(k_1x_1 + b_1)^2 + b_2 - y_1)k_2(k_1x_1 + b_1) \\
\checkmark\ e_{k_1} &= 4(k_2(k_1x_1 + b_1)^2 + b_2 - y_1)k_2(k_1x_1 + b_1)x_1 \\
\checkmark\ e_{b_1} &= 4(k_2(k_1x_1 + b_1)^2 + b_2 - y_1)k_2(k_1x_1 + b_1)
\end{aligned}
$$