CompSci 527 Midterm Exam Sample

The exam will be closed-book, closed-notes, and you will not be allowed to have anything other than the exam and a pen/pencil and an eraser on your desk. The amount of space provided under each question is *not* an indication of the length of the answer. Materials covered are all the materials in sections 0-3 of the class syllabus page and the first four homework assignments. Materials on the syllabus page include class notes and all the supplementary materials except those listed in parentheses. Appendices of the class notes are *not* required reading, except to the extent that they help you understand materials in the main text.

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.

$I =$

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

$H =$

$C =$

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.]

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

$H =$

2	3
1	1

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

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

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.

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.

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:

$$\begin{aligned} I(x,y) &= I(a,b) \quad \underline{\qquad\qquad} \\ &+ I(a+1,b) \quad \underline{\qquad\qquad} \\ &+ I(a,b+1) \quad \underline{\qquad\qquad} \\ &+ I(a+1,b+1) \quad \underline{\qquad\qquad} \end{aligned}$$

Fill in the missing parts above.

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

8. What are “hard examples” in Dalal and Triggs’s training methodology for HoG descriptors?

9. What is a categorical set?

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

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?

12. Briefly describe an outcome in dart-throwing that has low variance and 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 '$-$').

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?

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

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

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

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.