

CS 479/679 Pattern Recognition Sample Midterm Exam

1. **[20 pts]** Define the following terms (be brief and specific):

(a) Curse of dimensionality

(b) Generalization performance

(c) Conditional risk (or expected loss)

(d) Receiver Operating Characteristic (ROC) Curve

(e) Genetic algorithm

2. **[15 pts]** Consider a c -class classification problem; under what conditions would the optimal classifier be equivalent to (a) the minimum distance classifier? (b) the Mahalanobis distance classifier?

3. **[15 pts]** What is the goal of feature selection and why is it important? What are the main categories of search and evaluation methods for feature selection and how are they different?

4. **[15 pts]** Consider the following probability distribution:

$$p(x) = \begin{cases} (\theta + 1)x^\theta & \text{for } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Given n points x_1, x_2, \dots, x_n sampled from the above distribution, derive a formula for the maximum likelihood estimate of θ .

5. Consider the problem of dimensionality reduction.

(a) **[5 pts]** How are PCA and LDA different?

(b) **[15 pts]** Consider the eigenface approach; what are the steps for (i) face recognition and (ii) face detection?

6. [15 pts] Using the Bayesian network below, answer the following question: suppose all we know is that a fish is thin, has medium lightness, and was caught in south Atlantic; what season is it now, most likely?

