Mid Term Project Part 1

- Measurement Space $D = \times_{n=1}^{N} L_n$
  - Measurement d has N components $d = (d_1, \ldots, d_N)$
  - Input $|L_n| = M_n, \; n = 1, \ldots, N$
- Given $(d_1, \ldots, d_N)$ compute the linear address
- Input K, The Number of Classes
- Input e the $K \times K$ economic gain matrix
- Input $P(d \mid c)$ the class conditional probabilities
- Input $P(c)$ the prior probabilities
- Compute $P(c \mid d)$
- Compute the Discrete Bayes Rule $f_d(c)$
- Compute the $K \times K$ Confusion Matrix
- Compute the Expected Gain

Mid Term Project Part 2

- Discrete Bayes
- Class 0 Prior Probability .4
- Class 1 Prior Probability .6
- Use economic gain matrix

$$\begin{pmatrix} 1 & -1 \\ -2 & 3 \end{pmatrix}$$

- Use Data Set 1
  - Features f_1, f_2, f_3, f_4, f_5, $0 < f_i < 1$
  - Class Label 0,1
  - 10,000,000 Records
  - Each record 5 floating point numbers followed by 0 or 1
  - 544 MB file
- Use a memory M=10,000 for each class conditional probability

Optimize The Probability Estimation Parameters

- Split the ground truth sample into three parts
- Use the first part to calculate the quantizer and probabilities
- Calculate the decision rule
- Apply the decision rule to the second part so that an unbiased estimate of the expected economic gain given the decision rule can be computed
- Brute force optimization to find the values of smoothing parameter k to maximize the estimated expected gain
- With k fixed, use the third part to determine an estimate for the expected gain and confusion matrix for the optimization