1. (25 pts.) Show that if the likelihood function is Gaussian with unknown mean \( \mu \) and unknown covariance matrix \( \Sigma \), then the maximum likelihood estimates are given by

\[
\hat{\mu} = \frac{1}{N} \sum_{k=1}^{N} x_k
\]

and

\[
\hat{\Sigma} = \frac{1}{N} \sum_{k=1}^{N} (x_k - \hat{\mu})(x_k - \hat{\mu})^T
\]

where \( N \) is the number of samples, and \( A^T \) denotes the transpose of the vector \( A \).

2. (25 pts.) Consider the problem of classifying an instance \( x \) into one of two classes \( \omega_1 \) and \( \omega_2 \) with prior probabilities \( P(\omega_1) \) and \( P(\omega_2) \). Suppose the classification loss matrix

\[
L = \begin{bmatrix}
\lambda_{11} & \lambda_{12} \\
\lambda_{21} & \lambda_{22}
\end{bmatrix}
\]

Derive from first principles, the classification rule that minimizes the risk of misclassification.

3. (25 pts.) Define and state the key properties of well-behaved attribute evaluation functions for decision tree construction from training data. Give some examples of well-behaved attribute evaluation functions.

4. (25 pts.) Prove that the beta prior results in a beta posterior when estimating the parameters of the binomial distribution in the maximum a posteriori setting.

5. (25 pts.) Consider a sequence of labeled training instances \( \mathcal{E} \) of cars described by 4 attributes: origin, manufacturer, color, and model:

- (USA Dodge Red Luxury)
- (Japan Honda Blue Economy)
- (USA Dodge Red Sports)
- (Italy Fiat White Economy)
- (USA Ford Blue Economy)
- (USA Chrysler White Economy)
- (Japan Honda White Luxury)
- (USA Mercury Black Economy)
- (Japan Nissan Red Economy)
- (USA Dodge Red Luxury)
- (Korea Hyundai Black Economy)
- (Germany Mercedes Red Luxury)
Consider the class of concepts that can be expressed as decision trees. Construct a decision tree from the labeled examples. Show the calculations used to select the attributes to test at the top two levels of the decision tree using 2-way splits and entropy reduction as the splitting criterion. Repeat the calculations using Gini Index as the splitting criterion.

6. (25 pts.) Prove that the entropy $H(p_1 \cdots p_m)$ has a unique maximum at $p_1 = p_2 = \cdots = p_m = \frac{1}{m}$. (Hint: $\forall i \; p_i \geq 0$; $\sum_{i=1}^{m} p_i = 1$; and it helps to first show that $H$ is a convex function of $p_1 \cdots p_m$.)