Artificial Intelligence — Warm Up Test

1. (15 points) **Bayes' Rule**

   (a) (5 points) State Bayes' Rule and describe in one or two sentences what it is used for.

   (b) (5 points) A test for a rare disease has:

   - True positive rate: $P(\text{Positive} \mid \text{Disease}) = 0.98$
   - False positive rate: $P(\text{Positive} \mid \text{No Disease}) = 0.05$
   - Disease prevalence: $P(\text{Disease}) = 0.01$

   Compute the probability that a person who tests positive actually has the disease.

   (c) (5 points) Explain the concept of conditional independence and give one example.

2. (20 points) **Bayesian Networks**

   Consider the following Bayesian network structure:

$$A \to C, \quad B \to C, \quad C \to D.$$

   where A and B are C's parent, and C is D's parent.

   (a) (5 points) List all conditional probability tables (CPTs) required to fully specify the joint distribution.

   (b) (5 points) Using the chain rule for Bayesian networks, express $P(A, B, C, D)$ as a product of conditional probabilities.

   (c) (5 points) Are A and B independent? Are they conditionally independent given C? Explain briefly.

   (d) (5 points) Suppose you observe $D = d$. Describe qualitatively how this evidence might affect the posterior distributions of A and B.

3. (25 points) **Decision Tree Learning**

   You are given the following dataset:

Outlook	**Temperature**	**Windy**	**Play?**
Sunny	Hot	False	No
Sunny	Hot	True	No
Overcast	Hot	False	Yes
Rain	Mild	False	Yes
Rain	Cool	False	Yes
Rain	Cool	True	No
Overcast	Cool	True	Yes
Sunny	Mild	False	No
Sunny	Cool	False	Yes
Rain	Mild	True	Yes
Sunny	Mild	True	Yes
Overcast	Mild	False	Yes
Overcast	Hot	True	Yes
Rain	Mild	False	Yes

(a) (10 points) Compute the entropy of the target attribute *Play?*.

(b) (10 points) Compute the information gain for the attribute *Outlook*.

(c) (5 points) Based on your calculations, indicate whether *Outlook* would be selected as the root node.

4. (20 points) **Linear Regression**

(a) (5 points) State the hypothesis function for linear regression and describe the meaning of the parameters.

(b) (5 points) Write the cost function (mean squared error) for linear regression with one variable, and compute its gradient with respect to w_0 and w_1.

(c) (5 points) Explain the difference between *underfitting* and *overfitting*. Provide one example of each.

(d) (5 points) Consider the following dataset: (1, 2), (2, 2.5), (3, 3.7), (4, 4.2). Compute w_0 and w_1 in $h_{\mathbf{w}} = w_1 x + w_0$ to best fit the data.

5. (25 points) **Comprehensive Problem**

Consider a medical diagnosis system with the following Bayesian network:

$$\text{GeneticRisk} \rightarrow \text{Disease}, \qquad \text{Disease} \rightarrow \text{Symptom}.$$

Probabilities:

$$P(G = 1) = 0.2, \quad P(D = 1 \mid G = 1) = 0.3, \quad P(D = 1 \mid G = 0) = 0.05,$$

$$P(S = 1 \mid D = 1) = 0.9, \quad P(S = 1 \mid D = 0) = 0.1.$$

(a) (10 points) Compute $P(D = 1)$.

(b) (10 points) Compute $P(D = 1 \mid S = 1)$ using Bayes' Rule.

(c) (5 points) Briefly describe whether observing $S = 1$ increases, decreases, or does not change the probability of $G = 1$, and explain why.