Name:	Roll No:	Marks:

Data Mining and Machine Learning

Quiz 2, II Semester, 2025–2026

19 February, 2026

Some questions below may have more than one correct answer. You get full credit if you select all correct options. You get partial credit if you select a non-empty, strict subset of correct options. You get zero credit if you select any incorrect option.

1. We use regression to fit a degree four polynomial to inputs with two features (x_1, x_2). How many derived non-linear features do we have to add?

   (a) 5
   (b) 12 ✔
   (c) 16
   (d) 28

   Explanation:

   There are 3 quadratic features: $\{x_1^2, x_1x_2, x_2^2\}$.

   There are 4 degree three features: $\{x_1^3, x_1^2x_2, x_1x_x^2, x_2^3\}$.

   There are 5 degree four features: $\{x_1^4, x_1^3x_2, x_1^2x_2^2, x_1x_x^3, x_2^4\}$.

2. In logistic regression:

   (a) Squared error is not a good loss function because it does not follow from MLE.
   (b) We use cross entropy as the loss function to make gradient descent converge faster. ✔
   (c) We use the sigmoid function because the step function is not differentiable. ✔
   (d) The input to the sigmoid function can be any monotonically increasing function of the input attributes. ✔

   Explanation:

   (a) — a loss function need not be justified based on MLE. Squared error is bad because the rate of convergence of gradient descent could be slow.

3. In Bayesian classification:

   (a) The naïve Bayes assumption that attributes are independent accurately models the ground reality in many applications.
   (b) The naïve Bayes assumption helps build classifiers with much smaller volumes of training data. ✔
   (c) Laplace smoothing accounts for attribute values that are not observed in the training data. ✔
   (d) Laplace smoothing can be used to complete rows in the training data where some attribute values are missing.

   Explanation:

   (a) — The naïve Bayes assumption is typically *not* justified for most problems, but classifiers built on the assumption of indepedence often work well in practice.

   (d) — Laplace smoothing corrects for zero counts. It cannot be used to impute values for missing data.