Course: Machine Learning - Foundations
Week 2 - Practice Questions

1. Consider a function $f(x)$ such that

$$f(x) = \begin{cases} \dfrac{\sin(x)}{x} & , \ x \neq 0 \\ 1 & , \ x = 0 \end{cases}$$

Is $f(x)$ continuous at $x = 0$?

A. False

B. True

Answer: B

Solution:

$$f(x) = \begin{cases} \dfrac{\sin(x)}{x} & , \ x \neq 0 \\ 1 & , \ x = 0 \end{cases}$$

Now, f is continuous at $x = 0$ if $\lim\limits_{x \to a^+} f(x) = \lim\limits_{x \to a^-} f(x) = f(a)$.

First we will compute the left hand limit at $x = 0$:

$$\begin{aligned} LHL &= \lim_{x \to 0^-} f(x) \\ &= \lim_{h \to 0} f(0 - h) \\ &= \lim_{h \to 0} \frac{sin(-h)}{-h} \\ &= \lim_{h \to 0} \frac{sin(h)}{h} \\ &= 1 \end{aligned}$$

Next we will compute the right hand limit at $x = 0$:

$$\begin{aligned} RHL &= \lim_{x \to 0^+} f(x) \\ &= \lim_{h \to 0} f(0 + h) \\ &= \lim_{h \to 0} \frac{sin(h)}{h} \\ &= 1 \end{aligned}$$

$\because LHL = RHL = f(0) \implies f$ is continuous at $x = 0$

Since the statement is true, option B is correct.

2. If $U = [10, 100]$, $A = [30, 50]$ and $B = [50, 90]$, which of the following is/are false? (Consider all values to be integers.)

   A. $A^C = [10, 30] \cup [50, 100]$
   B. $A^C = [10, 30) \cup (50, 100]$
   C. $A \cup B = [30, 90]$
   D. $A \cap B = \emptyset$
   E. $A \cap B = \{50\}$
   F. $A^C \cap B^C = [10, 30) \cup [91, 100]$

Answer: A, D, F

Solution:
We know that,

$$\begin{aligned} A^c &= U \setminus A \\ &= [10, 100] \setminus [30, 50] \\ &= [10, 30) \cup (50, 100] \end{aligned}$$

$\therefore$ Option A is false and option B is true.

Next,

$$\begin{aligned} A \cup B = & \ [30, 50] \cup [50, 90] \\ = & \ [30, 90] \end{aligned}$$

$\therefore$ Option C is true.

Next,

$$\begin{aligned} A \cap B = & \ [30, 50] \cap [50, 90] \\ = & \ \{50\} \neq \emptyset \end{aligned}$$

$\therefore$ Option D is false and option E is true

Next,

$$\begin{aligned} A^c \cap B^c = & \ (A \cup B)^c \\ = & \ U \setminus (A \cup B) \\ = & \ [10, 100] \setminus [30, 90] \\ = & \ [10,\ 30) \cup (90, 100] \\ = & \ [10,\ 30) \cup [91, 100] \end{aligned}$$

Finally, option F is True.

3. Consider two d-dimensional vectors $\mathbf{x}$ and $\mathbf{y}$ and the following terms:
(i) $\mathbf{x}^T\mathbf{y}$
(ii) $\mathbf{x}.\mathbf{y}$
(iii) $\sum_{i=1}^{d} x_i y_i$
Which of the above terms are equivalent?

    A. Only (i) and (ii)

    B. Only (ii) and (iii)

    C. Only (i) and (iii)

    D. (i), (ii) and (iii)

Answer: D

Solution:

We have $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_d \end{bmatrix}$,

Now, $\mathbf{x}^T\mathbf{y} \;=\; \begin{bmatrix} x_1 & x_2 & \dots & x_d \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_d \end{bmatrix} \;=\; \displaystyle\sum_{i=1}^{d} x_i y_i$

Also by definition, $\mathbf{x} \cdot \mathbf{y} \;=\; \displaystyle\sum_{i=1}^{d} x_i y_i$

Therefore, all three terms are equivalent.

$\therefore$ Option D is correct.

4. The linear approximation of $tan(x)$ around $x = 0$ is:

    A. $1 + x$

    B. $1 - x$

    C. $x - 1$

    D. x

Answer: D

Solution:
The linear approximation of a function f around $x = a$ is given by

$$L(x) \;=\; f(a) \;+\; (x - a)f'(a)$$

Here, $f(x) = \tan(x)$ and $a = 0$, Substituting these in the above equation, we get

$$\begin{aligned} L(x) = &\ tan(0) \ + \ (x-0)\ sec^2(0) \qquad [\because (tanx)' = sec^2 x] \\ = &\ x \end{aligned}$$

$\therefore$ Option D is correct

5. The partial derivative of $x^3 + y^2$ w.r.t. x at $x = 1$ and $y = 2$ is __________.

Answer: 3

Solution:

The partial derivative of $x^3 + y^2$ w.r.t. x is

$$\frac{\partial}{\partial x}\left(x^3 + y^2\right) = \ 3x^2$$

$\therefore$ At $(x, y) = (1, 2)$, it will be

$$\frac{\partial}{\partial x}\left(x^3 + y^2\right)\Big|_{(x,y)=(1,2)} = 3$$

So, the answer is 3.

6. Consider the following function:

$$f(x) = \begin{cases} 7x + 2, & \text{if } x > 1 \\ 9, & \text{if } x \leq 1 \end{cases}$$

Is $f(x)$ continuous?

A. Yes

B. No

Answer: A

Solution:
Now, $\forall\, x < 1$, f is a linear function which makes it continuous and $\forall\, x \geq 1$, f is constant which also makes it continuous. So, the only point we need to check is at $x = 1$:

$$\begin{aligned} LHL = &\ \lim_{x \to 1^-} f(x) \\ = &\ \lim_{h \to 0} f(1-h) \\ = &\ \lim_{h \to 0} 9 \\ = &\ 9 \end{aligned}$$

$$
\begin{aligned}
RHL &= \lim_{x \to 1^+} f(x) \\
&= \lim_{h \to 0} f(1+h) \\
&= \lim_{h \to 0} 7(1+h) + 2 \\
&= 9
\end{aligned}
$$

$\therefore LHL = RHL = f(1) = 9 \implies f$ is continuous at $x = 1 \implies f$ is continuous.
So, option A is correct.

7. Which of the following is the best approximation of $e^{0.019}$? (Use linear approximation around 0).

    A. 1

    B. 0

    C. 0.019

    D. 1.019

Answer: D

Solution:
The linear approximation of a function f around $x = a$ is given by

$$L(x) \ = \ f(a) \ + \ (x-a)f'(a)$$

Here, $f(x) = e^x$ and $a = 0$, substituting these values in $L(x)$, we get

$$
\begin{aligned}
L(x) &= \ e^0 \ + \ (x-0)e^0 \qquad [\because \ (e^x)' = e^x] \\
&= \ 1 + x
\end{aligned}
$$

$\implies e^{0.019} \ \approx \ L(0.019) \ = \ 1 + 0.019 \ = \ 1.019.$
$\therefore$ Option D is correct.

8. What is the linear approximation of $f(x, y) = x^2 + y^2$ around (1, 1)?

    A. $2x + 2y + 2$

    B. $2x + 2y - 2$

    C. $2x + 2y + 1$

    D. $2x + 2y - 1$

Answer: B

Solution:
The linear approximation of a function f around $(x, y) = (a, b)$ is given by

$$
\begin{aligned}
L(x, y) &= f(a, b) + \begin{bmatrix} x - a \\ y - b \end{bmatrix} \cdot \nabla f(a, b) \\
&= f(a, b) + (x - a)\ f_x(a, b) + (y - b)\ f_y(a, b)
\end{aligned}
$$

Here, $(a, b) = (1, 1)$ and $f(x, y) = x^2 + y^2$,

$$
\nabla f(x, y) = \begin{bmatrix} 2x \\ 2y \end{bmatrix} \implies \nabla f(1, 1) = \begin{bmatrix} 2 \\ 2 \end{bmatrix}
$$

So the linear approximation is given by

$$
\begin{aligned}
L(x, y) &= f(1, 1) + (x - 1)(2) + (y - 1)(2) \\
&= 2x + 2y - 2
\end{aligned}
$$

$\therefore$ Option B is correct.

9. What is the gradient of $f(x, y) = x^2 y$ at (1, 3)?
   A. [1, 6]
   B. [6, 1]
   C. [1, 3]
   D. [3, 1]

Answer: B

Solution:
The gradient (∇f) of $f(x, y) = x^2 y$ is

$$
\begin{aligned}
\nabla f(x, y) &= \begin{bmatrix} f_x(x, y) \\ f_y(x, y) \end{bmatrix} \\
&= \begin{bmatrix} 2xy \\ x^2 \end{bmatrix}
\end{aligned}
$$

$\therefore$ At $(x, y) = (1, 3)$, it will be

$$
\nabla f(1, 3) = \begin{bmatrix} 6 \\ 1 \end{bmatrix}
$$

So, option B is correct.

10. The directional derivative of $f(x, y, z) = x^2 + 3y + z^2$ at (1, 2, 1) along the unit vector in the direction of [1, -2, 1] is __________.

Answer: -0.816

Solution:
The directional derivative is given by

$$D_{\hat{\mathbf{u}}} f(1,2,1) \;=\; \nabla f(1,2,1) \cdot \hat{\mathbf{u}}$$

First, let us compute the gradient

$$\nabla f(x,y,z) = \begin{bmatrix} f_x \\ f_y \\ f_z \end{bmatrix} = \begin{bmatrix} 2x \\ 3 \\ 2z \end{bmatrix}$$

$$\nabla f(1,2,1) = \begin{bmatrix} 2 \\ 3 \\ 2 \end{bmatrix}$$

The unit vector in the direction of $\mathbf{u} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$ is $\hat{\mathbf{u}} = \frac{\mathbf{u}}{||\mathbf{u}||} = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$.

$$\therefore\; D_{\hat{\mathbf{u}}} f(1,2,1) \;=\; \nabla f(1,2,1) \cdot \hat{\mathbf{u}} \;=\; \begin{bmatrix} 2 \\ 3 \\ 2 \end{bmatrix} \cdot \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \;=\; \frac{-2}{\sqrt{6}} \;\approx\; -0.816.$$

So, the answer is -0.816.

11. Find the direction of steepest ascent for the function $x^2 + y^3 + z^4$ at point (1, 1, 1).
    A. $\left[\frac{2}{\sqrt{29}}, \;\frac{3}{\sqrt{29}}, \;\frac{4}{\sqrt{29}}\right]$
    B. $\left[\frac{-2}{\sqrt{29}}, \;\frac{3}{\sqrt{29}}, \;\frac{4}{\sqrt{29}}\right]$
    C. $\left[\frac{-2}{\sqrt{29}}, \;\frac{-3}{\sqrt{29}}, \;\frac{4}{\sqrt{29}}\right]$
    D. $\left[\frac{2}{\sqrt{29}}, \;\frac{-3}{\sqrt{29}}, \;\frac{4}{\sqrt{29}}\right]$

Answer: A

Solution:
The direction of the steepest ascent for any function is the direction of the gradient itself.

So let us compute the gradient first

$$\nabla f(x, y, z) = \begin{bmatrix} f_x \\ f_y \\ f_z \end{bmatrix}$$

$$= \begin{bmatrix} 2x \\ 3y \\ 4z \end{bmatrix}$$

$$\nabla f(1, 1, 1) = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$$

The direction of the steepest ascent, $\hat{\mathbf{u}}$ is in the direction of the gradient. That is,

$$\hat{\mathbf{u}} = \frac{\nabla f(1,1,1)}{||\nabla f(1,1,1)||} = \frac{1}{\sqrt{29}} \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}.$$

$\therefore$ Option A is correct.

12. The directional derivative of $f(x, y, z) = x + y + z$ at point (-1, 1, -1) along the unit vector in the direction of [1, -1, 1] is ________.

Answer: 0.577

Solution:

The directional derivative is given by

$$D_{\hat{\mathbf{u}}} f(-1, 1, -1) = \nabla f(-1, 1, -1) \cdot \hat{\mathbf{u}}$$

First, let us compute the gradient

$$\nabla f(x, y, z) = \begin{bmatrix} f_x \\ f_y \\ f_z \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\nabla f(-1, 1, -1) = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

The unit vector in the direction of $\mathbf{u} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$ is $\hat{\mathbf{u}} = \frac{\mathbf{u}}{||\mathbf{u}||} = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$

$$\therefore \; D_{\hat{\mathbf{u}}} f(-1, 1, -1) = \nabla f(-1, 1, -1) \cdot \hat{\mathbf{u}} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \cdot \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{3}} \approx 0.577$$

So, the answer is 0.577.

13. Which of the following is/are the vector equations of a line that passes through (1, 2, 3) and (4, 0, 1)?

(i) $[x, y, z] = [1, 2, 3] + \alpha[3, -2, -2]$
(ii) $[x, y, z] = [4, 0, 1] + \alpha[-3, 2, 2]$
(iii) $[x, y, z] = [1, 2, 3] + \alpha[4, 0, 1]$
(iv) $[x, y, z] = [4, 0, 1] + \alpha[1, 2, 3]$

A. (i) and (ii)
B. (iii) and (iv)
C. (i) and (iii)
D. (ii) and (iv)

Answer: E

Solution:
The vector equation of a line passing through two points **a** and **b** is given by

$$[x, y, z] \ = \ \mathbf{a} \ + \ \alpha \ (\mathbf{b} - \mathbf{a}) \ \ , \ \alpha \in \mathbb{R}$$

Taking $\mathbf{a} = [1, 2, 3]$ and $\mathbf{b} = [4, 0, 1]$, we have

$$\begin{aligned} [x, y, z] &= \ [1, 2, 3] \ + \ \alpha \ (\ [4, 0, 1] - [1, 2, 3] \ ) \\ &= \ [1, 2, 3] \ + \ \alpha \ [3, -2, -2] \end{aligned}$$

Now, taking $\mathbf{a} = [4, 0, 1]$ and $\mathbf{b} = [1, 2, 3]$, we have

$$\begin{aligned} [x, y, z] &= \ [4, 0, 1] \ + \ \alpha \ (\ [1, 2, 3] - [4, 0, 1] \ ) \\ &= \ [4, 0, 1] \ + \ \alpha \ [-3, 2, 2] \end{aligned}$$

So, statements (i) and (ii) are correct.

14. As per Cauchy-Schwarz inequality, if **b** is a -ve scalar multiple of **a**, then,

A. $\mathbf{a}^T\mathbf{b} \geq ||\mathbf{a}|| * ||\mathbf{b}||$
B. $\mathbf{a}^T\mathbf{b} \leq ||\mathbf{a}|| * ||\mathbf{b}||$
C. $\mathbf{a}^T\mathbf{b} = ||\mathbf{a}|| * ||\mathbf{b}||$
D. $\mathbf{a}^T\mathbf{b} = -||\mathbf{a}|| * ||\mathbf{b}||$

Answer: D

Solution:
The Cauchy-Schwarz inequality states

$$|\mathbf{a}^T\mathbf{b}| \leqslant ||\mathbf{a}|| * ||\mathbf{b}||$$

$$\implies -||\mathbf{a}|| * ||\mathbf{b}|| \leqslant \mathbf{a}^T\mathbf{b} \leqslant ||\mathbf{a}|| * ||\mathbf{b}||$$

Also, equality holds if and only if in the boundary case when **a** is a scalar multiple of **b**, since here **a** and **b** are -ve scalar multiples, we get

$$-||\mathbf{a}|| * ||\mathbf{b}|| = \mathbf{a}^T\mathbf{b}$$

$\therefore$ Option D is correct.

Course: Machine Learning - Foundations
Week 1 (Practice questions)

---

1. (1 point) Which of the following is true about a model?

    1. A model is a mathematical representation of reality.
    2. A model is an exact representation of a system.
    3. A model uses no assumptions.

        A. 1 and 2

        B. 2 and 3

        C. only 1

        D. only 2

    Answer: C

    Explanation: A model represents reality/a system in some mathematical form or another.
    A model is never an exact representation, it is a mathematical simplification of reality.
    A model always makes some assumptions. For instance, this assumption could be, that the data points fit in 2 dimensions or 3 dimensions.

---

2. (1 point) Identify which of the following problem requires regression algorithm.

    A. Predicting the country that a person belongs to based on his physical features.

    B. Predicting the animal present in a given image based on sample images of various animals.

    C. Predicting the topic of a given Wikipedia article based on article's keywords.

    D. Predicting the stock price of a company on a given day based on revenue growth and profit after tax.

    Answer: D

    Explanation: We know that there are 2 broad categories for supervised algorithms: regression and classification.
    Classification is where we have to put the data points into some "class" based on the features and regression is where we predict a numerical value. In A, B & C, we would classify the data points into some class(es) whereas, option D requires us to predict a numerical value.

---

3. (1 point) For spam detection, if we use traditional programming rather than a machine learning approach, which problems may be faced?

A. The need to check for a long list of possible patterns that may be present.

B. Difficulty in maintaining the program containing the complex hard-coded rules.

C. The need to keep writing new rules as the spammers become innovative.

D. All of these.

Answer: D

Explanation: Traditional programming would require explicitly defining and checking for specific patterns that indicate spam. This would involve creating a long list of possible patterns that may be present in spam messages.
Traditional programming for spam detection often involves creating complex rules and conditions to identify spam. Maintaining and updating such a program can be cumbersome.
If you compare the spam messages from a decade ago, you'll notice a heavy change in the pattern. So, traditional programming will require us to constantly update our code and rules to keep up with the spam detection trend.

---

4. (1 point) In a regression model, the parameters w_i's and b of the function $f(X) = \sum_{j=1}^{d} w_j x_j + b$ where $X = [x_1, x_2, ..., x_d]$

   A. are strictly integers.

   B. always lie in the range [0,1].

   C. are any real value.

   D. are any imaginary value.

Answer: C

Explanation: The weights w and the bias term b is not restricted to integers or a specific range of real numbers. They can be of any real value. Regression models don't make use of imaginary numbers.

---

5. (1 point) Identify classification problem in the following statements

   A. Find the gender of a person by analyzing writing style.

   B. Predict the price of a car based on engine power, mileage etc.

   C. Predict whether there will be abnormally heavy rainfall tomorrow or not based on previous data.

   D. Predict the number of street accidents that will happen this month based on traffic volume count.

Answer: A,C

If the given problem statement allows you to predict the results into a specific class from a specific set of classes, then it can be said that the problem is a classification problem.

In option A, based on the writing style, the data point would be tagged as a specific class - Either “Male”, “Female” or any other class, if present.

In option B, price is a numerical quantity that is being predicted. Hence, this would be a regression problem, rather than a classification problem.

In option C, there are 2 classes for which the predictions will be made. The classes can be, for instance, “will_rain” and “will_not_rain”

In option D, the prediction is a numerical quantity. Therefore, this will be a regression problem

---

6. (1 point) Identify task that needs the use of regression in the following

   A. Predict the height of a person based on his weight.

   B. Predict the country a person belongs to based on his linguistic features.

   C. Predict whether the price of gold will increase tomorrow or not based on data of last 25 day.

   D. Predict whether a movie is comedy or tragedy based on its reviews.

Answer: A

Explanation: For regression problems, the predicted value must be a numerical quantity. This numerical quantity is usually continuous.

In option A, you are predicting a numerical value. Therefore, this will fall under the category of regression problem.

In option B, specific classes are predicted. In this case, the classes would be: the set of all countries

In option C, again, the predictions being made are of specific classes and not numerical quantities. Here, classes are: { will_increase, will_not_increase }

In option D, the specific style/genre of the movies is to be predicted. Here, the classes would be: { Comedy, tragedy }

---

7. (1 point) Which of the following are examples of unsupervised learning problems?

   A. Grouping tweets based on topic similarity

   B. Making clusters of cells having similar appearance under microscope.

   C. Checking whether an email is spam or not.

   D. Identify the gender of online customers based on buying behaviour.

Answer: A,B

Explanation: To identify unsupervised learning problems, see if the problem requires us to group/cluster based on the patterns in the data. Unsupervised learning problems have unlabelled data (the target column is absent).

In option A, we are grouping, hence unsupervised.

In option B we are making clusters, hence unsupervised.

In option C, We typically train a model for the following problem using labelled data to help us identify spam or not spam for future unseen emails (based on the patterns learned from the labelled data). Hence, it's a supervised problem.

In option D, we would be "grouping" the customers based on their buying behaviour to help us identify their gender. We don't have explicit labels available with us. So, this is also an unsupervised problem.

---

8. (1 point) Which of the following is/are incorrect?

   A. $\mathbf{1}(2\ \textit{is even}) = 1$

   B. $\mathbf{1}(10\%3 = 0) = 0$

   C. $\mathbf{1}(0.5 \notin \mathbb{R}) = 0$

   D. $\mathbf{1}(2 \in \{2, 3, 4\}) = 0$

Answer: D

Explanation: First, we need to understand the notation used here.
1(expression) = output $\rightarrow$ if the expression inside () is true, then the output is 1, if the expression inside () is false, the output is 0.

This is known as the indicator function. In mathematics, an indicator function of a subset of a set is a function that maps elements of the subset to one, and all other elements to zero.

Now, in option A, the expression is true, hence the output will be 1. Therefore, it is a correct option.

In option B, the expression is false (10 mod 3 is equal to 1), hence the output is 0. Therefore, this option is correct as well.

In option C, the expression is false, hence the output should be 0. Again, the option is correct.

In option D, the expression is true but the output mentioned is 0, whereas it should have been 1. Therefore, it's an incorrect statement and hence, our answer.

---

9. (1 point) Which of the following functions corresponds to a classification model?

A. $f : \mathbb{R}^d \to \mathbb{R}$

B. $f : \mathbb{R}^d \to \{+1, -1\}$

C. $f : \mathbb{R}^d \to \mathbb{R}^{d'}$ where $d'<d$

Answer: B

Explanation: Only option B is where we have a set of classes. In classification models, a set of features is mapped to a set of classes.

---

10. (1 point) Which of the following can form a good encoder decoder pair for d dimensional data ?

A. $f : \mathbb{R}^d \to \mathbb{R}^{d'}$ where $d' < d$
$g : \mathbb{R}^{d'} \to \mathbb{R}^d$ where $d' > d$

B. $f : \mathbb{R}^d \to \mathbb{R}^{d'}$ where $d' > d$
$g : \mathbb{R}^{d'} \to \mathbb{R}^d$ where $d' < d$

C. $f : \mathbb{R}^d \to \mathbb{R}^{d'}$ where $d' < d$
$g : \mathbb{R}^{d'} \to \mathbb{R}^d$ where $d' < d$

D. $f : \mathbb{R}^d \to \mathbb{R}^{d'}$ where $d' > d$
$f : \mathbb{R}^{d'} \to \mathbb{R}^d$ where $d' > d$

Answer: C

Explanation: A good encoder would be the one which reduces the dimensions.
A good decoder would be one which gives us higher dimensional data (preferably the same dimension as the original data).
Only option C satisfies the aforementioned requirements.

---

11. (2 points) Consider the following two scenarios:
(1) Given the details of a person's sample, the lab technician wants to find whether the person is suffering from cancer or not.
(2) You are the manufacturer of a mobile company. You wish to know how many of the mobiles are expected to be sold in the next six months.

A. Both (1) and (2) are suited for regression.

B. Both (1) and (2) are suited for classification.

C. Problem (1) is better suited for classification while (2) is better suited for regression.

D. Problem (2) is better suited for classification while (1) is better suited for regression.

Answer: C

Explanation: In scenario 1, we have 2 classes: cancer and not-cancer. This scenario is suitable for classification. In scenario 2, we have a regression problem as we want to predict the number of mobiles to be sold. Therefore, option C is the correct answer.

---

12. (2 points) Consider the following data set where each data point consists of three features x_1, x_2 and x_3:

x_1	x_2	x_3
10	10	9
13	12	13
5	5	4
8	7	7

Consider two encoder functions f and $\tilde{f}$ with decoders g and $\tilde{g}$ respectively aiming to reduce the dimensionality of the data set from 3 to 1:
Pair 1:$f(x_1, x_2, x_3) = x_1 - x_2 + x_3$ and $g(u) = [u, u, u]$
Pair 2:$\tilde{f}(x_1, x_2, x_3) = \dfrac{x_1 + x_2 + x_3}{3}$ and $\tilde{g}(u) = [u, u, u]$
The reconstruction loss of the encoder decoder pair is the mean of the squared distance between the reconstructed input and input.

Answer: Pair 1: 3(Range 2.95 to 3.05)
Pair 2: 0.667(Range 0.63 to 0.7)

Explanation:

We can calculate the $f()$ and $\tilde{f}()$ as following:

x_1	x_2	x_3	$f(x_1,\ x_2,\ x_3)$	$\tilde{f}(x_1,\ x_2,\ x_3)$
10	10	9	9	9.66
13	12	13	14	12.66
5	5	4	4	4.66
8	7	7	8	7.33

Now, we can also calculate the $g(u)$ and $\tilde{g}(u)$ which would give us the vectors

$g(u)$	$\tilde{g}(u)$
[9, 9, 9]	[9.66, 9.66, 9.66]
[14, 14, 14]	[12.66, 12.66, 12.66]
[4, 4, 4]	[4.66, 4.66, 4.66]
[8, 8, 8]	[7.33, 7.33, 7.33]

Now, we calculate the loss using the loss function below:

$$loss_{pair1} = \frac{1}{n}\sum_{i=0}^{n} ||g\left(f\left(x^i\right)\right) - x^i||^2$$

$$= \frac{||[-1,-1,0]||^2 + ||[1,2,1]||^2 + ||[-1,-1,0]||^2 + ||[0,1,1]||^2}{4}$$

$$= \frac{2+6+2+2}{4} = 3$$

$$loss_{pair2} = \frac{1}{n}\sum_{i=0}^{n} ||\ g\left(f\left(x^i\right)\right) - x^i||^2$$

$$= \frac{||[0.34,\ 0.34, 0.66]||^2 + ||[0.34, 0.66, 0.34]]||^2 + ||[0.34, 0.34, 0.66]||^2 + ||[0.67, 0.33, 0.33]||^2}{4}$$

$$= \frac{0.6668 + 0.6668 + 0.6668 + 0.66674}{4} = 0.6667$$