Krishnaraj Thadesar
PA 10
Batch A1

classmate
Date
Page

AI Theory Assignment 2

Q.1: Explain how does gradient descent work in Linear Regression?

→ Gradient descent is an optimization algorithm commonly used in machine learning to minimize the cost function during the training of a model. In the context of linear regression, the goal is to find the optimal parameters of a linear equation that best fits the given data.

→ Its objective is to minimize the cost function, which measures the difference between predicted and actual values.

* **Model**: Linear Regression Equation.

$$y = mx + b.$$

* **Cost Function**: Typically, Mean Squared Error (MSE)

* **Parameters**: Slope (m) and Intercept (b)

* **Algorithm**:
  1. Initialize m & b.
  2. Iterate until convergence

> Compute partial derivatives of the cost function with respect to m and b.

> Update m and b using derivatives & a learning rate (α)

* Update Rules:

$$m = m - \alpha \frac{\partial J}{\partial m} \qquad b = b - \alpha \frac{\partial J}{\partial b}$$

* Partial Derivatives (MSE)

$$\frac{\partial J}{\partial m} = -\frac{1}{m} \sum_{i=1}^{m} x_i (m x_i + b)$$

$$\frac{\partial J}{\partial b} = -\frac{1}{m} \sum_{i=1}^{m} (y_i - m x_i + b)$$

* Convergence:

→ Stop when the change is the cost function is small.

* Result

→ Optimal parameters m & b for the best linear fit to the data

Q2. Name some Evaluation Metrics for Regression Model ~~for~~ & when you would use one?

→ Several evaluation metrics are commonly used to assess the performance of regression models. The choice for a specific technique depends on the nature of the data and the goals of the analysis.

→ Here are some common regression techniques.

① Mean Absolute Error (MAE)

Formula : $$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$$

Use case : MAE is suitable when the absolute ~~errors~~ errors are important and you want a metric that is not sensitive to outliers.

② Mean Squared Error : (MCE)

Formula : $$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Use case : MSE is widely used and emphasis larger errors due to the squaring effect; however, it is sensitive to outliers.

③ Root Mean Squared Error (RMSE):

Formula : $$RMSE = \sqrt{MSE}$$

Use case : Similar to MSE but the square root is taken to bring the metric back to the same scale as the target variables. It is also sensitive to outliers.

④ **Mean Absolute Percentage Error** (MAPE)

Formula : $MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100$

Use case: Useful when you want to express errors as a percentage of the actual values and you want a metric that is easy to intercept

⑤ R- squared (R^2) :

Formula : ~~RSS~~ ~~TSS~~ ε

$$R^2 = 1 - \sum_{i=1}^{n} \frac{(Y_i - \hat{Y}_i)^2}{(Y_i - \bar{Y})^2}$$

Use case: Measures the proportion of the variance in the dependent variable that is predictable from the independent variable.

R^2 values range from 0 to 1, with higher values indicating a better fit.

⑥ **Adjusted R - squared**:

Formula : $\text{Adjusted } R^2 = 1 - \frac{(1-R^2)(n-1)}{n-p-1}$

where, n = no. of observations
p = no. of predictions.

* **Use case**: Similar to R^2, but adjusted for the no of predictors. It penalize the inclusion of irrelevant variables.

7) Mean Squared Logarithmic Error (MSLE)

Formula: $$MSLE = \frac{1}{n}\sum_{i=1}^{n}\left(\log(1+Y_i) - \log(1+\hat{Y}_i)\right)^2$$

Use case: Useful when you want to penalize underestimates more than overestimates; and the target variable has a wide range.

The choice of the evaluation metric depends on the specific goals of the analysis, the characteristics of the data, & the consequences of different types of errors in predictions. It's often a good idea to use a combination of metrics to get a comprehensive understanding of model performance.

classmate

Date
Page

Q.5.

	Softmax function	vs. Sigmoid Function
1.	Used for multi classification in logistic regression model	Used for binary classification in logistic regression model.
2.	The probability sum will be 1	The probability sum need not be 1
3.	Used in different layers of neural networks	Used as activation function while building neural networks
4.	The high value will have the higher probability than others	The high value will have high probability but not the highest probability.
5.	Formula : $\sigma(x) = \frac{1}{1 - e^{-x}}$	$softmax(x)_i = \frac{e^{x_i}}{\sum_{j=1}^{k} e^{x_j}}$

Q. (4) Explain the terms overfitting and underfitting

Ans: Overfitting :

Def: The model learns the training data too well, capturing noise instead of the underlying pattern.

characteristics :

→ Performs well on training data but poorly on new unseen data.

→ Memorises specific data points rather than general trends.

→ Too complex, capturing noise

Causes :

→ Using a model that is too complex
→ Training for too many epochs, allowing the model to memorize the training data.
→ Having insufficient data leading the model to fit noise rather than actual patterns.

Prevention & Remedies :

→ Using simpler models or regularization techniques to reduce complexity.

→ Increase the amount of training data.

→ Employ techniques like cross-validation to assess model performance on unseen data.

eg:

→ Training a polynomial regression model with a high degree on a small dataset fitting the noise instead of the underlying trend.

Underfitting:

Def: This occurs when the model is too simple to capture the underlying pattern of the data.

Characteristics

→ The model performs poorly on both new and unseen data.

→ It fails to capture the complexity of the underlying patterns in the data.

Causes:

→ Using the model that is too simple.

→ ~~In~~ Insufficient training, when the model hasn't learnt the underlying patterns.

Prevention & Remedies.

→ Use a more complex model or increase

model capacity:

→ Provide more relevant features to the model

→ Train for more epochs to allow the model to learn from the data.

eg: Using a linear regression model on a dataset with a non-linear relationship, resulting is a poor fit.

23/1/2023