Samiksha N. Dalvi 2310180004 INFT B

ML Open Book Test

1\. Explain the difference between Z-score Standardization & Min-Max scaling in terms of formula, effect & use cases.

→ Z score Standardization :-

Formula = $x' = \frac{x-\mu}{\sigma}$

where μ is the mean & σ is the standard deviation

Effect: Data is transformed to have mean = 0 & sd = 1. This ensures all features contribute equally, regardless of their original units.

Use Cases: Works well when the data follows a normal distribution. Preferred in algorithms like logistic Regression, SVM, PCA, Linear Regression where assumptions about data distribution exist.

Min Max Scaling :

Formula : $x' = \frac{x - x_{min}}{x_{max} - x_{min}}$

Effect: Rescales features into a fixed range, typically [0, 1] (can also be [-1, 1]). It preserves the relationships b/w values but compresses everything to a bounded range.

Use Cases: Preferred for algorithms that rely on distance calculations (K-NN, K-Means) or neural networks where bounded inputs speed up gradient descent.

Date :
Page :

2. For an SVM, the kernel function (like RBF) is sensitive to feature scales. Since the feature scales. Since the features already follow a Gaussian distribution, the most suitable method is Z-score Standardization
Reason: Z score ensures each feature has mean 0 & varience 1, which matches the Gaussian assumption & prevents 1 feature from domination due to scale.

3. In this case, outliers will distort both mean & min-max scaling.
Best method: Robust Scaling

$$x' = \frac{x - \text{median}}{IQR}$$

Reason: Robust scaling uses the median instead of mean & spreads values according to IQR. Outliers do not heavily affect the median or IQR, making this method more stable.

4. The transformation $x' = \frac{x-\mu}{\sigma}$

Name: Z-score standardization
Effect: Produces data with mean = 0 & standard deviation = 1
When not suitable:-
- If the dataset has outliers, the mean & σ become distorted. Example: a single huge outlier will inflate σ & compress all other values.
- In such cases robust scaling is preferred.

5 b. The tanh function outputs values in [-1, 1]
- If inputs are scaled using Min-Max into [-1, 1] they match the activation's range → leads to faster training & convergence.
- If Z-score scaling is used, values may fall outside [-1, 1] causing neurons to saturate at extreme values.

6. Decision Trees & Random Forests:-
→ Work by splitting data based on feature thresholds
→ Thresholds are independent of feature magnitude.
Hence scaling doesn't affect splits.

K-NN :-
→ Relies on distance metrics like Euclidean distance
→ Example:- If one feature is in km's & another in m's, the feature with larger scale will dominate
→ Hence, scaling (Z-score or Min-max) is essential.

7. • Cosine similarity measures the angle between 2 vectors not their magnitude.
- Therefore absolute scaling is not important but L2 normalization is often used to ensure all document vectors have unit length.
- Eg. Doc A = [2, 2, 2], Doc B = [100, 100, 100] both represent the same direction. Without normalization they might differ, but cosine similarity treats them equally after normalization

8.

Method	Formula	Effect on Mean/Var	Suitability for outliers
1. Z-score Standardization	$\frac{(x-\mu)}{\sigma}$	Mean = 0, SD = 1	Sensitive to outliers
2. Min-Max Scaling	$\frac{(x-x_{min})}{(x_{max}-x_{min})}$	Range [0,1] or [-1,1]	Very sensitive
3. Robust scaling	$\frac{(x-\text{median})}{IQR}$	Median-centered, scale = IQR	Robust to outliers

9. • Problem: outliers cause the range (max - min) to become too large.

→ Eg. Suppose data is mostly 10-20, but one value = 1000. After Min-Max scaling, 10-20 compresses into [0, 0.01] while 1000 maps to 1.

→ this makes normal values indistinguishable.

• Better choice: Robust scaling (median, IQR) or Z-score with outlier clipping/ winsorization.

10. • Problem: Gradient descent updates depend on scale. If one feature has values in millions & another in decimals, the larger scaled features ~~domain~~ dominates. This causes:-

→ slow covergence or

→ Oscillations in optimization.

Fix:- 1) Zscore Standardization: brings all features to mean 0 & var 1

2) Min Max scaling: compresses features to [0,1]

3) Both methods make features comparable & ensure ... gradient updates.