

ASSIGNMENT-2

Q1

Overfitting:

happens when model performs very well on the training data but fails to generalise on new/unseen data.

Scenarios:

- model is too complex
- Training data is smaller or noisy
- Model trained for too many epochs
- No regularisation or dropout applied.

Underfitting:

It happens when a model is too simple to learn the underlying patterns in data.

Scenarios

- model has high bias
- Insufficient training or poor feature selection.
- Too much regularization used
- Model not trained ~~enough~~ for enough epochs.

Q2

Bias-Variance Trade off:

- balance b/w bias (error due to simplifying) & variance (error due to sensitivity).

High bias : leads to underfitting

High variance : leads to overfitting

SPPU-BE-COMP-CONTENT - KSKA Git

Goal is to find a model with low bias & low variance for best performance.

Q3

Error (residual) for each sample:

$$e_i = y_i - \hat{y}_i$$

Total error measured using Mean Squared Error (MSE)

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

The model learns coefficients that minimise this MSE value

Q4

Ensemble Learning:

combines multiple models to create a stronger and more accurate model

Types

1. Bagging: Trains models in parallel to reduce variance (Random Forest)
2. Boosting: Trains models sequentially to reduce bias (Ada Boost)
3. Stacking: combines outputs of various models using a meta model.

SPPU-BE-COMP-CONTENT - KSKA Git

Q5

ADABOOST WORKING:

1. Start with equal weights for all samples.
2. Train a weak model (decision stump)
3. Increase weights of misclassified samples
4. Train the next model focusing on these samples
5. Combine all models using weighted voting

Final model = weighted sum of all weak learners.

Q6

- Accuracy: $(TP + TN) / (TP + TN + FP + FN)$
- Precision: $TP / (TP + FP)$ → correctness of positive predictions.
- Recall Sensitivity: $TP / (TP + FN)$ → ability to find positives
- F1 Score: $2 \times (Precision \times Recall) / (Precision + Recall)$
- Roc-Auc: Measures model's ability to distinguish b/w classes
- Confusion Matrix: Table showing TP, FP, TN, FN to analyze performance.