|  |  |
| --- | --- |
| **Overfitting** | **Underfitting** |
| The model is too complex, memorizes the training data, and performs poorly on unseen data. | The model is too simple and cannot capture the underlying patterns in the data, leading to poor performance on both training and test datasets. |
| Very low training error,  Low bias, High Variance | High training error,  High bias, Low Variance |
| Poor generalization to new data (the model learns noise). | Poor generalization due to oversimplification (the model misses important patterns). |
| Too many features or model parameters, lack of regularization, poor hyperparameters selection or too much training on noisy data. | Too few features or overly simplistic assumptions, lack of sufficient training data or model flexibility. |
| A polynomial regression model of high degree that fits the training data perfectly but performs poorly on new data. | A linear regression model used for a non-linear problem, failing to capture the data’s trends. |

The goal is to find a **balance** fit where the model is complex enough to capture the underlying patterns but simple enough to generalize well to new data. This balance is achieved through techniques like **cross-validation**, **regularization**, and selecting the appropriate model complexity.

=========================================

**Reasons for overfitting:**

Poor Complex Model, Hyperparameters Selection, poor feature selection (not very relevant features for predicting target), insufficient training data., inadequate validation (k-fold cross validation), lack of regularization.

**Reasons for underfitting:**

Simple model, insufficient training data, insufficient features (noth-east corridor, closer to train station, house facing), Insufficient training time(epochs), inadequate validation (after many splits, o/p not consistent), too much of regularisation

==============