

Def;

The phenomenon where a model learns the distribution of the training samples too closely, rather than the actual underlying distribution

Overfitting can lead to several problems:

- **Poor generalization to unseen data:** The model performs very well on the training data but performs poorly on new, unseen data (validation or test data). This is the core problem. The model has learned the noise and specific details of the training data, rather than the underlying patterns.
- **High variance:** The model is very sensitive to small changes in the training data. If you train the model on a slightly different training set, you might get a very different model and different performance.
- **Unrealistic model complexity:** Overfit models tend to be more complex than necessary. They might have a large number of parameters or highly complex relationships, which makes them difficult to interpret and understand.
- **Wasted resources:** Training an overfit model can be a waste of time and computational resources, as the resulting model is not useful for real-world applications.

overfitting leads to a model that is good at memorizing the training data but bad at making predictions on new data. The goal of machine learning is to build models that generalize well, and overfitting prevents that.