Choosing the right ML model

Created	@July 31, 2025
Edited	@August 2, 2025 7:31 AM
Archive	☐

Model selection in ML is the process of choosing the best suited model for a particular problem. Selecting a model depends on various factors such as dataset, task, nature of the model etc.

The key challenge lies in finding the balance between model complexity and performance. Some factors to consider when selecting a model include:

- Data size and quality: Complex models require more data to avoid overfitting
- Interpretability needs: Some applications require transparent decision-making
- Computational resources: Training and inference costs vary significantly between models
- Performance metrics: Different evaluation metrics may be appropriate depending on the problem

Models can be selected based on:

1. Type of Data Available

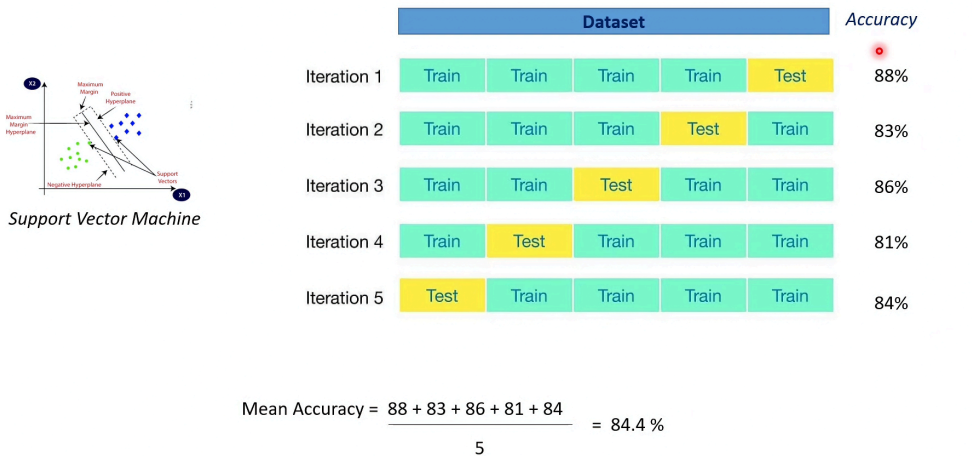
- Images and Videos - Convolutional Neural Networks (CNN)
- Text and Speech - Recurrent Neural Networks (RNN)
- Numerical Data - Support Vector Machines (SVM), Logistic Regression, Decision Trees

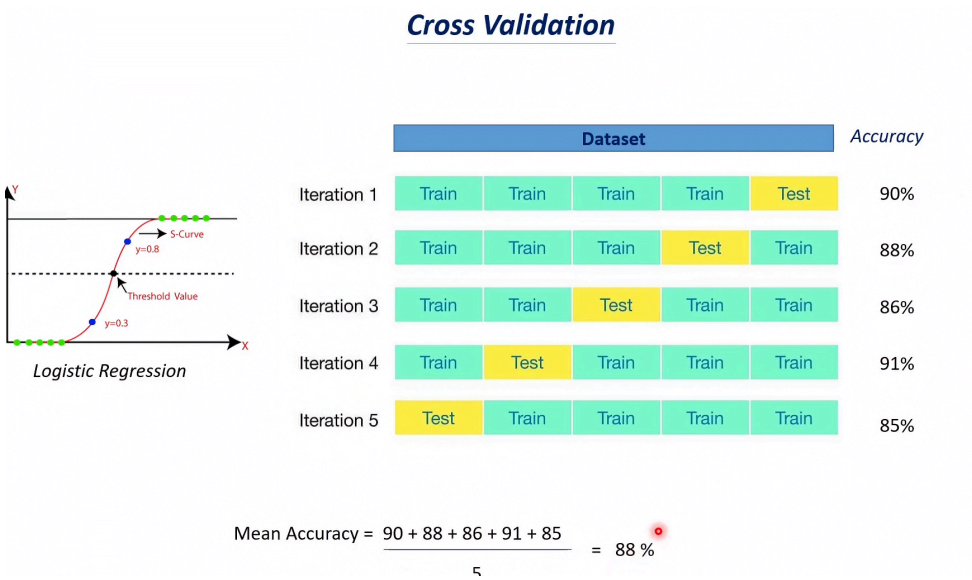
2. Based on the Task

- Classification Tasks - SVM, Logistic Regression, Decision Trees
- Regression Tasks - Linear Regression, Random Forest, Polynomial Regression
- Clustering Tasks - K-means Clustering, Hierarchical Clustering

Cross Validation:

based on the accuracy, we choose the right model





accuracy score of logistic regression is greater, so it is the best fit model for this problem

Implementation:

```
>>> from sklearn import datasets, linear_model
>>> from sklearn.model_selection import cross_val_score
>>> diabetes = datasets.load_diabetes()
>>> X = diabetes.data[:150]
>>> y = diabetes.target[:150]
>>> lasso = linear_model.Lasso()
>>> print(cross_val_score(lasso, X, y, cv=3))
[0.33150734 0.08022311 0.03531764]
```