# SUPERVISED MACHINE LEARNING

# CLASSIFICATION

- Classification is a process of finding a function which helps in dividing the dataset into classes based on different parameters.
- In Classification, a computer program is trained on the training dataset and based on that training, it categorizes the data into different classes.
- The task of the classification algorithm is to find the mapping function to map the input(x) to the discrete output(y).
- The best example to understand the Classification problem is Email Spam Detection. The model is trained based on millions of emails on different parameters, and whenever it receives a new email, it identifies whether the email is spam or not. If the email is spam, then it is moved to the Spam folder.

# BASIC TERMINOLOGY IN CLASSIFICATION ALGORITHMS

| Terminology | Description |
|---|---|
| Classifier | An algorithm that maps the input data to a specific category. |
| Classification model | A classification model tries to draw some conclusion from the input values given for training. It will predict the class labels/categories for the new data. |
| Feature | A feature is an individual measurable property of a phenomenon being observed. |
| Binary Classification | Classification task with two possible outcomes. Eg: Gender classification (Male / Female) |
| Multi-class classification | Classification with more than two classes. In multi-class classification, each sample is assigned to one and only one target label. Eg: An animal can be a cat or dog but not both at the same time. |
| Multi-label classification | Classification task where each sample is mapped to a set of target labels (more than one class). Eg: A news article can be about sports, a person, and location at the same time. |

# APPLICATIONS OF CLASSIFICATION ALGORITHMS

- Email spam classification
- Bank customers loan pay willingness prediction.
- Cancer tumor cells identification.
- Sentiment analysis
- Drugs classification
- Facial key points detection
- Pedestrians detection in an automotive car driving.

# TYPES OF CLASSIFICATION ALGORITHMS

- Logistic Regression
- K-Nearest Neighbours
- Support Vector Machines
- Kernel SVM
- Naïve Bayes
- Decision Tree Classification
- Random Forest Classification

# REGRESSION

- Regression is a process of finding the correlations between dependent and independent variables. It helps in predicting the continuous variables such as prediction of Market Trends, prediction of House prices, etc.
- The task of the Regression algorithm is to find the mapping function to map the input variable(x) to the continuous output variable(y).
- Suppose we want to do weather forecasting, so for this, we will use the Regression algorithm. In weather prediction, the model is trained on the past data, and once the training is completed, it can easily predict the weather for future days.

# TYPES OF REGRESSION ALGORITHMS

- Simple Linear Regression
- Multiple Linear Regression
- Polynomial Regression
- Support Vector Regression
- Decision Tree Regression
- Random Forest Regression

# DIFFERENCE BETWEEN REGRESSION AND CLASSIFICATION

| Regression Algorithm | Classification Algorithm |
|---|---|
| In Regression, the output variable must be of continuous nature or real value. | In Classification, the output variable must be a discrete value. |
| The task of the regression algorithm is to map the input value (x) with the continuous output variable(y). | The task of the classification algorithm is to map the input value(x) with the discrete output variable(y). |
| Regression Algorithms are used with continuous data. | Classification Algorithms are used with discrete data. |
| Regression algorithms can be used to solve the regression problems such as Weather Prediction, House price prediction, etc. | Classification Algorithms can be used to solve classification problems such as Identification of spam emails, Speech Recognition, Identification of cancer cells, etc. |
| The regression Algorithm can be further divided into Linear and Non-linear Regression. | The Classification algorithms can be divided into Binary Classifier and Multi-class Classifier. |

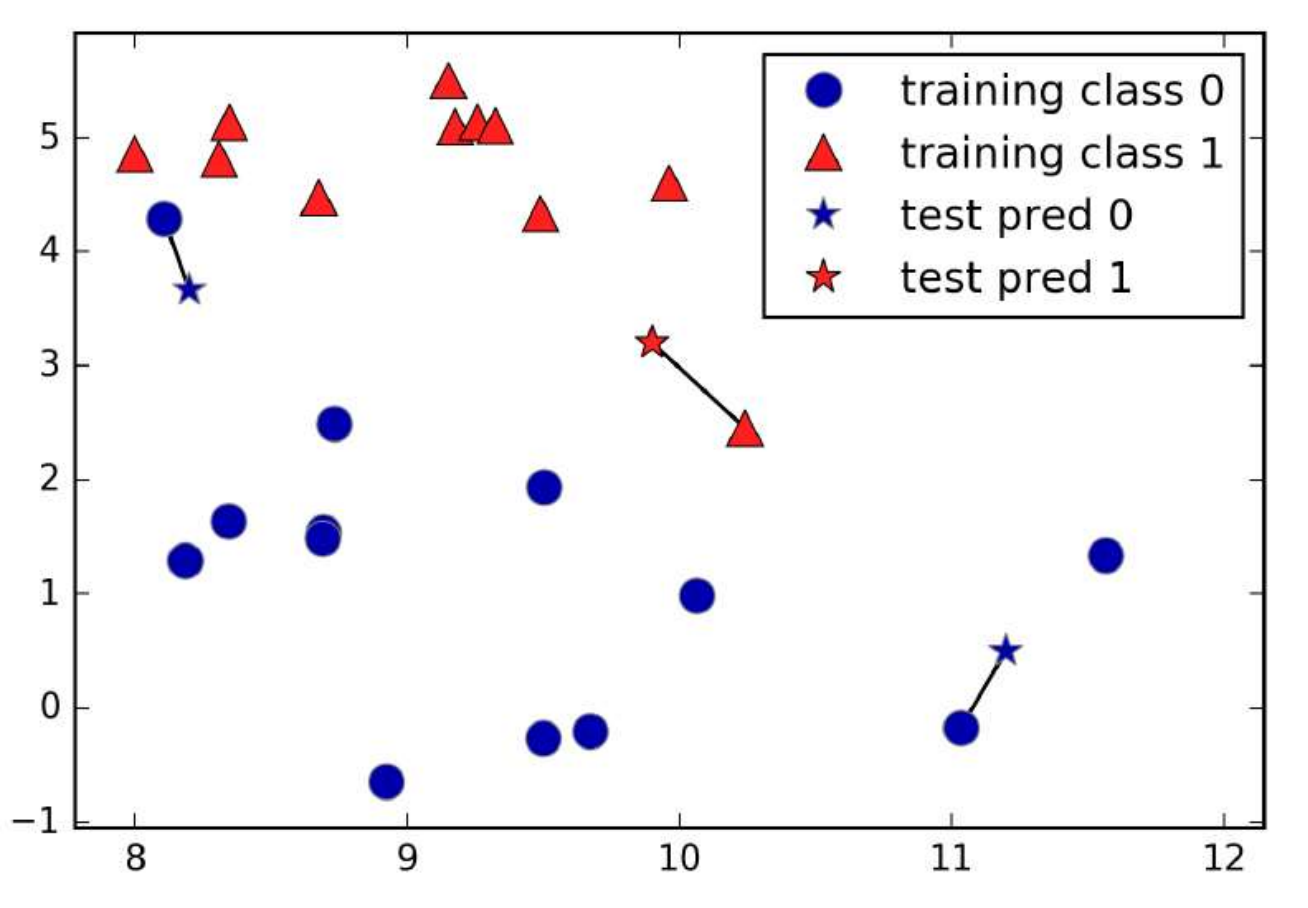
# SUPERVISED LEARNING ALGORITHMS

# K-NEAREST NEIGHBORS

- The k-NN algorithm is arguably the simplest machine learning algorithm.
- Building the model consists only of storing the training dataset.
- To make a prediction for a new data point, the algorithm finds the closest data points in the training dataset—its "nearest neighbors."

# K-NEIGHBORS CLASSIFICATION

- In its simplest version, the k-NN algorithm only considers exactly one nearest neighbor, which is the closest training data point to the point we want to make a prediction for.
- The prediction is then simply the known output for this training point.


training class 0
training class 1
test pred 0
test pred 1
5
4
3
2
1
0
−1
8
9
10
11
12

# K-NEIGHBORS CLASSIFICATION

- Instead of considering only the closest neighbor, we can also consider an arbitrary number, k, of neighbors. This is where the name of the k-nearest neighbors algorithm comes from.
- When considering more than one neighbor, we use voting to assign a label.
- This means that for each test point, we count how many neighbors belong to class 0 and how many neighbors belong to class 1.
- We then assign the class that is more frequent: in other words, the majority class among the k-nearest neighbors.

# K-NEAREST NEIGHBORS ALGORITHM USING SCIKIT-LEARN

ACTIVITY

# LINEAR MODELS

- Linear models are a class of models that are widely used in practice and have been studied extensively in the last few decades, with roots going back over a hundred years.
- Linear models make a prediction using a linear function of the input features.

# LINEAR MODELS FOR REGRESSION

- For regression, the general prediction formula for a linear model looks as follows:

  $\hat{y} = w[0] * x[0] + w[1] * x[1] + \ldots + w[p] * x[p] + b$

- For a dataset with a single feature, this is:

  $\hat{y} = w[0] * x[0] + b$

which you might remember from high school mathematics as the equation for a line.

Here, w[0] is the slope and b is the y-axis offset.

# LINEAR MODELS FOR CLASSIFICATION

- Linear models are also extensively used for classification.
- Let's look at binary classification first. In this case, a prediction is made using the following formula:

$$\hat{y} = w[0] * x[0] + w[1] * x[1] + \ldots + w[p] * x[p] + b > 0$$

The formula looks very similar to the one for linear regression, but instead of just returning the weighted sum of the features, we threshold the predicted value at zero.

If the function is smaller than zero, we predict the class –1; if it is larger than zero, we predict the class +1. This prediction rule is common to all linear models for classification.

Again, there are many ways to find the coefficients (w) and the intercept (b).

# NAÏVE BAYES CLASSIFIERS

- Naive Bayes classifiers are a family of classifiers that are quite like the linear models discussed in the previous section.
- However, they tend to be even faster in training. The price paid for this efficiency is that naive Bayes models often provide generalization performance that is slightly worse than that of linear classifiers like LogisticRegression and LinearSVC.
- The reason that naive Bayes models are so efficient is that they learn parameters by looking at each feature individually and collect simple per-class statistics from each feature.

# TYPES OF NAÏVE BAYES CLASSIFIERS

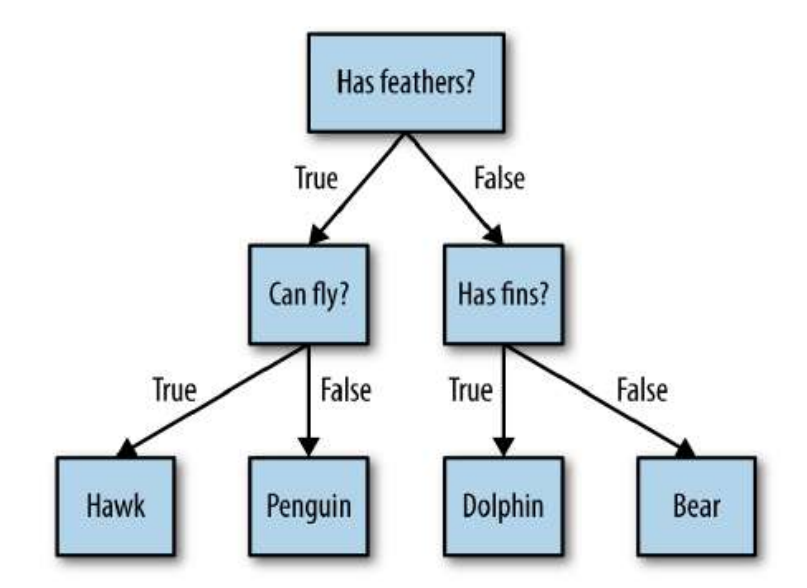
- There are three kinds of naive Bayes classifiers implemented in scikit-learn: GaussianNB, BernoulliNB, and MultinomialNB.
- GaussianNB can be applied to any continuous data, while BernoulliNB assumes binary data and MultinomialNB assumes count data (that is, that each feature represents an integer count of something, like how often a word appears in a sentence). BernoulliNB and MultinomialNB are mostly used in text data classification.
- The BernoulliNB classifier counts how often every feature of each class is not zero.

# STRENGTHS, WEAKNESSES, AND PARAMETERS

- The naive Bayes models share many of the strengths and weaknesses of the linear models. They are very fast to train and to predict, and the training procedure is easy to understand.
- The models work very well with high-dimensional sparse data and are relatively robust to the parameters.
- Naive Bayes models are great baseline models and are often used on very large datasets, where training even a linear model might take too long.

# DECISION TREES

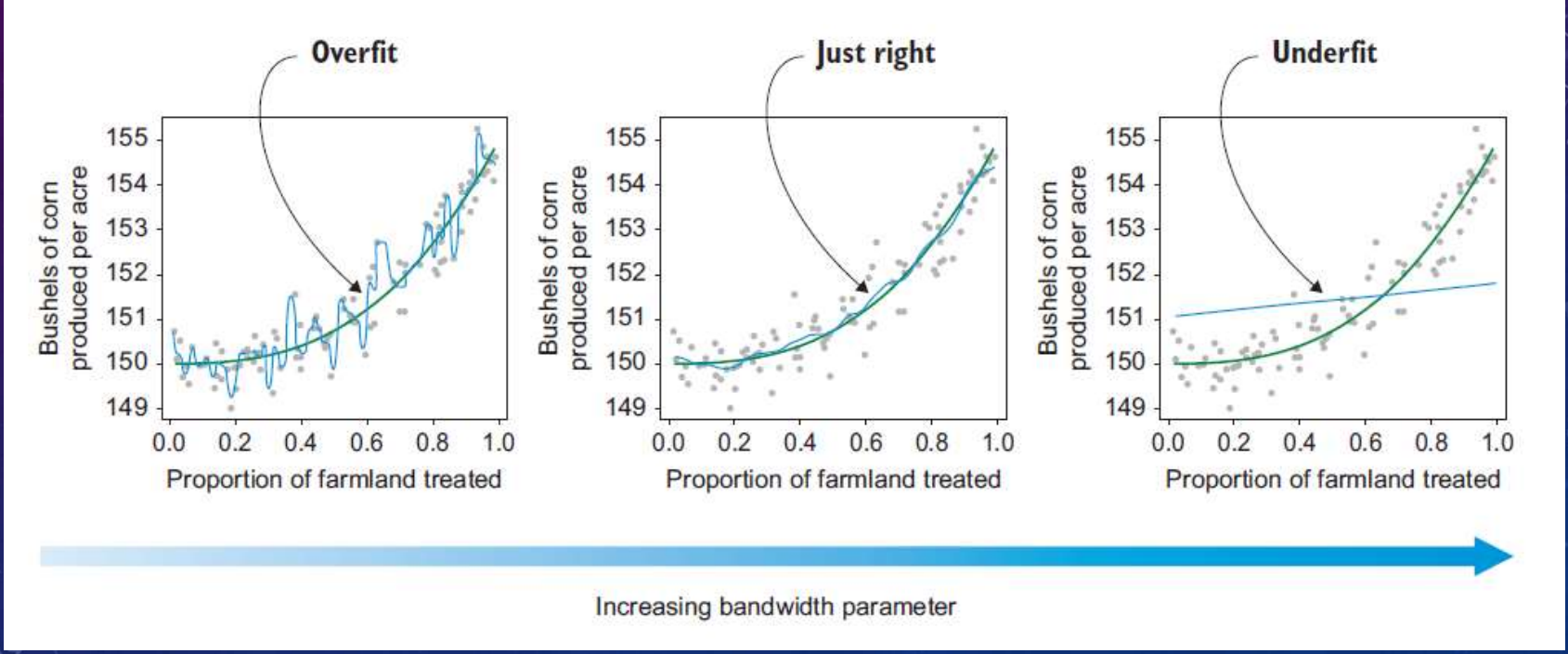
- Decision trees are widely used models for classification and regression tasks.
- Essentially, they learn a hierarchy of if/else questions, leading to a decision.



# MODEL EVALUATION AND GENERALIZATION

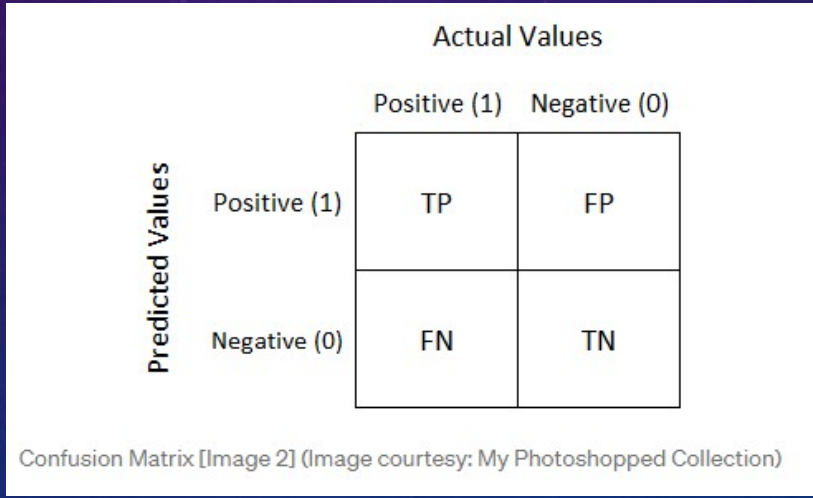
- The primary goal of supervised machine learning is accurate prediction. You want your ML model to be as accurate as possible when predicting on new data (for which the target variable is unknown).
- Said differently, you want your model, which has been built from training data, to generalize well to new data. That way, when you deploy the model in production, you can be assured that the predictions generated are of high quality.
- Therefore, when you evaluate the performance of a model, you want to determine how well that model will perform on new data.

# OVERFITTING AND MODEL OPTIMISM



# CONFUSION MATRIX

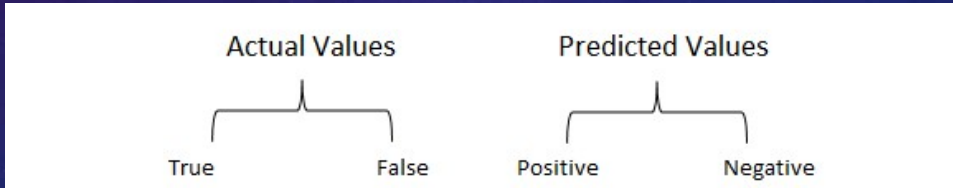
- It is a performance measurement for machine learning classification problem where output can be two or more classes. It is a table with 4 different combinations of predicted and actual values.




Confusion Matrix [Image 2] (Image courtesy: My Photoshopped Collection)

# TP, FP, FN, TN

- True Positive: You predicted positive and it's true.
- True Negative: You predicted negative and it's true.
- False Positive: (Type 1 Error) You predicted positive and it's false.
- False Negative: (Type 2 Error) You predicted negative and it's false.




Actual vs Predicted values [Image 4] (Image courtesy: My Photoshopped Collection)

# PRECISION, RECALL, ACCURACY, F-MEASURE

- Precision quantifies the number of positive class predictions that belong to the positive class. Precision should be high as possible.
- Recall quantifies the number of positive class predictions made from all positive examples in the dataset. Recall should be high as possible.
- Accuracy is the most intuitive performance measure, and it is simply a ratio of correctly predicted observation to the total observations. Accuracy should be high as possible.
- F-Measure provides a single score that balances both the concerns of precision and recall in one number.

# THE END