Module: Machine Learning

Live Session-3

Agenda:
Linear Regression
Logistic Regression
Implementations
Face Recognizer

Basic Regression Model

The Regression Problem

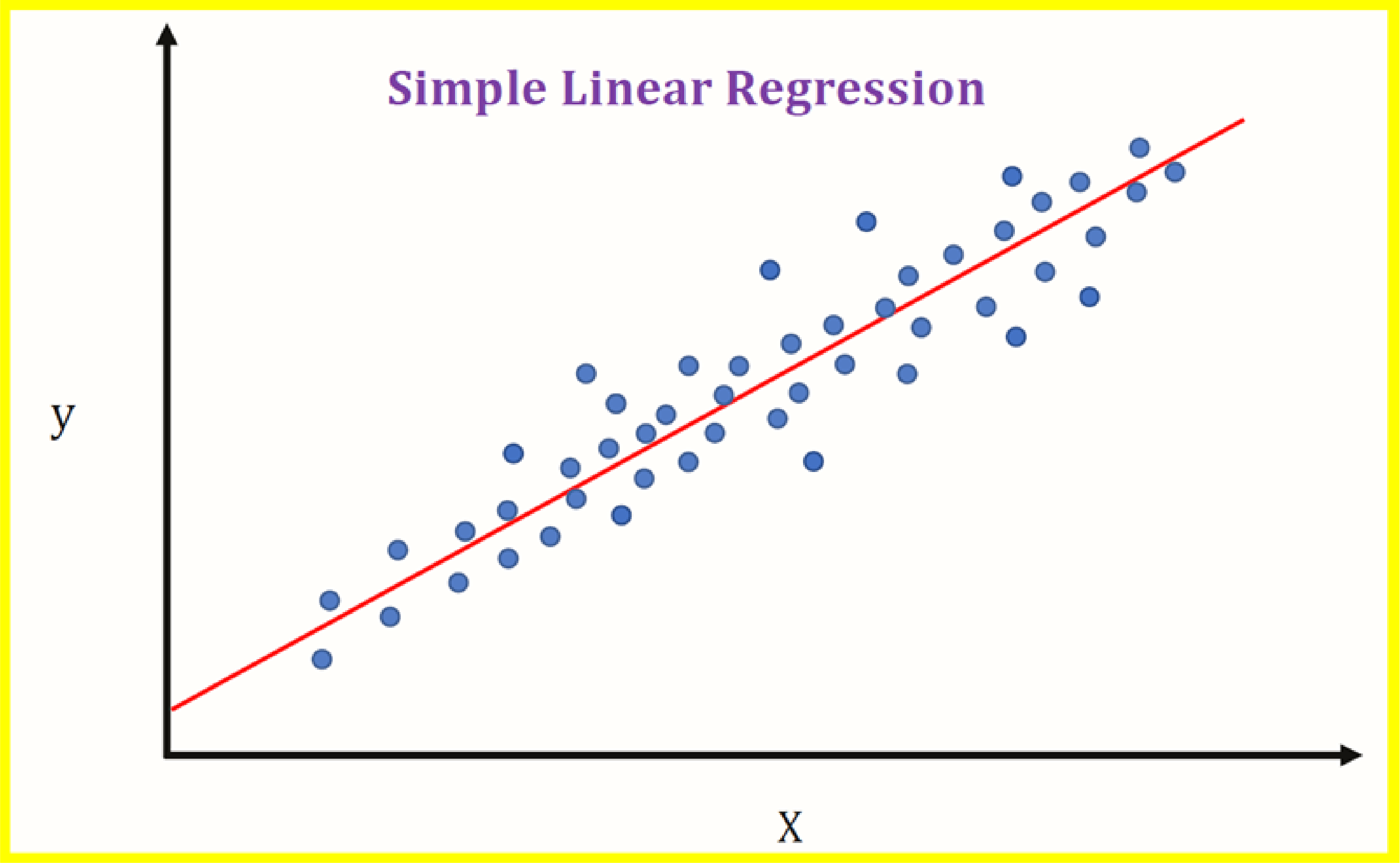
Regression refers to the problem of learning the relationships between some (qualitative or quantitative) input variables

$$X = [x_1, x_2, \dots, x_n]^T$$

and a quantitative output variable y. Means, regression is about learning a model f such that

$$y = f(X) + \epsilon$$

Simple Linear Regression



Multiple/Linear Regression Model

The linear regression model describes the output variable y (a scalar) as an affine combination of the input variables $x_1;\ x_2; \dots ; x_p$ (each a scalar) plus a noise term ϵ, i.e.

$$y = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_p x_p + \epsilon$$

where, a_i are the parameters of the regression model which we needs to estimate based on the given data.

How to learn the parameters $a_0, a_1, \dots, a_p$ from some training dataset $T = \{X_i, y_i\}$ for $i = 1,2,\dots, n$. Once, we estimate these parameters, future outputs for inputs that we have not yet seen can be predicted.

Linear Regression: Matrix Approach

Linear Regression: Calculus Approach

Mean Squared Error (MSE):

The residual for the i^{th} observation is given by:

$$r_i = y_i - \hat{y}_i = y_i - (\hat{a}_0 + \hat{a}_1 x_{i1} + \cdots + \hat{a}_p x_{ip})$$

The mean squared error (MSE) tells you how close a regression line is to a set of points. It does this by taking the distances from the points to the regression line (these distances are the “errors”) and squaring them. The squaring is necessary to remove any negative signs.

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

Goodness of fitting: Residual

Consider the linear regression models as $Y = \alpha + \beta X$ and $Y = \alpha$, then R^2 is defined as

$$R^2 = 1 - \frac{\text{Sum of square residuals from model with } \alpha \text{ and } \beta}{\text{Sum of square residuals from model with } \alpha \text{ only}}$$

Goodness of fitting

- We have $0 \leq R^2 \leq 1$
- If $SS(Res) = SS(total)$, then $R^2 = 0 \rightarrow$ model is not useful.
- If $SS(Res) = 0$, then $R^2 = 1 \rightarrow$ model fits all the points perfectly.

Essentially the same thing happens when there is more than one independent variables

How large does R^2 need to be to be considered as "good"?

This depends on the context, there is no absolute answer here. For hard to predict Y variables, smaller values may be "good".

Model Validation

One often attempts to "validate" a model either by:

- Randomly splitting an existing data set into two parts, and using part of the data for "model fitting", and part of the data for "model validation".

- Using one full data set for "model fitting", and finding a second independent data set for "model validation".

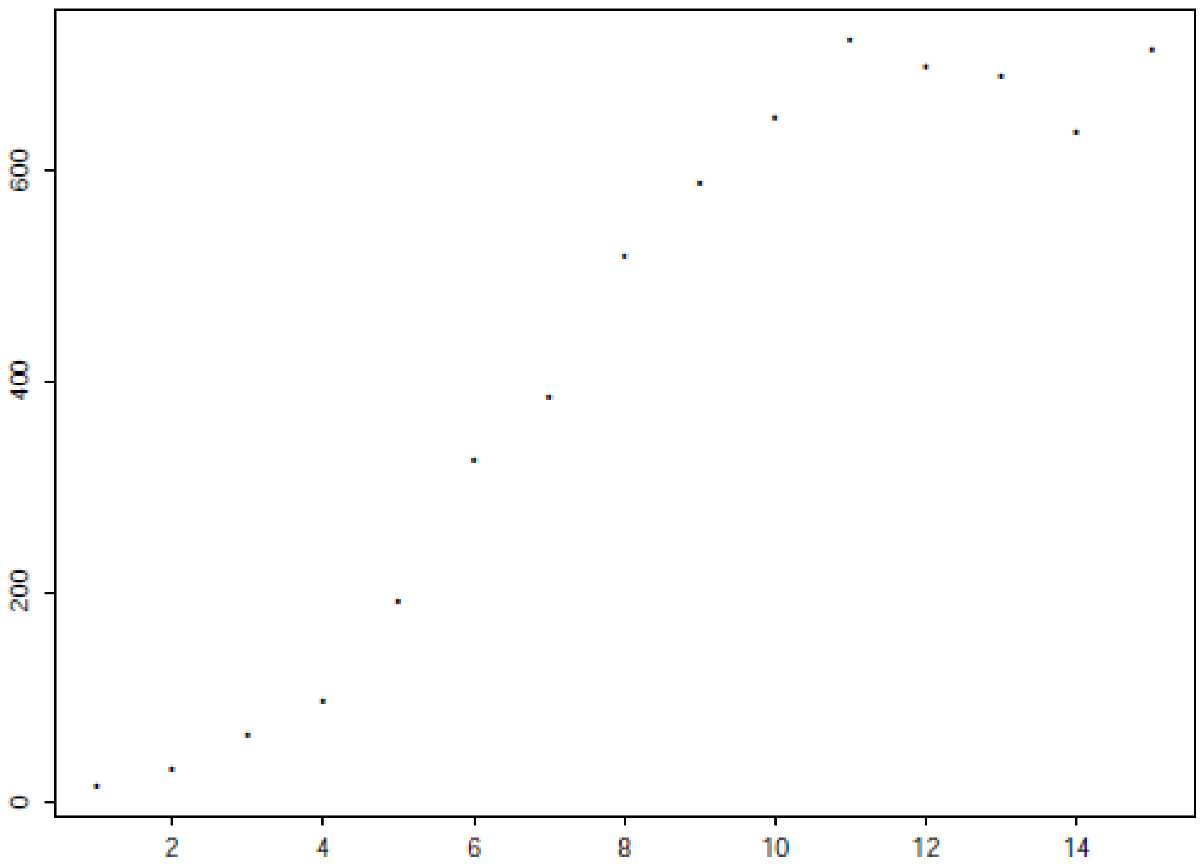
Polynomial Regression

Consider that the following table display data on the dry weight (Y) of 15 onion bulbs randomly assigned to 15 growing times (X) until measurement.

Growing Time	Dry Weight	Growing Time	Dry Weight
1	16.08	9	590.03
2	33.83	10	651.92
3	65.8	11	724.93
4	97.2	12	699.56
5	191.55	13	689.96
6	326.20	14	637.56
7	386.87	15	717.41
8	520.53		

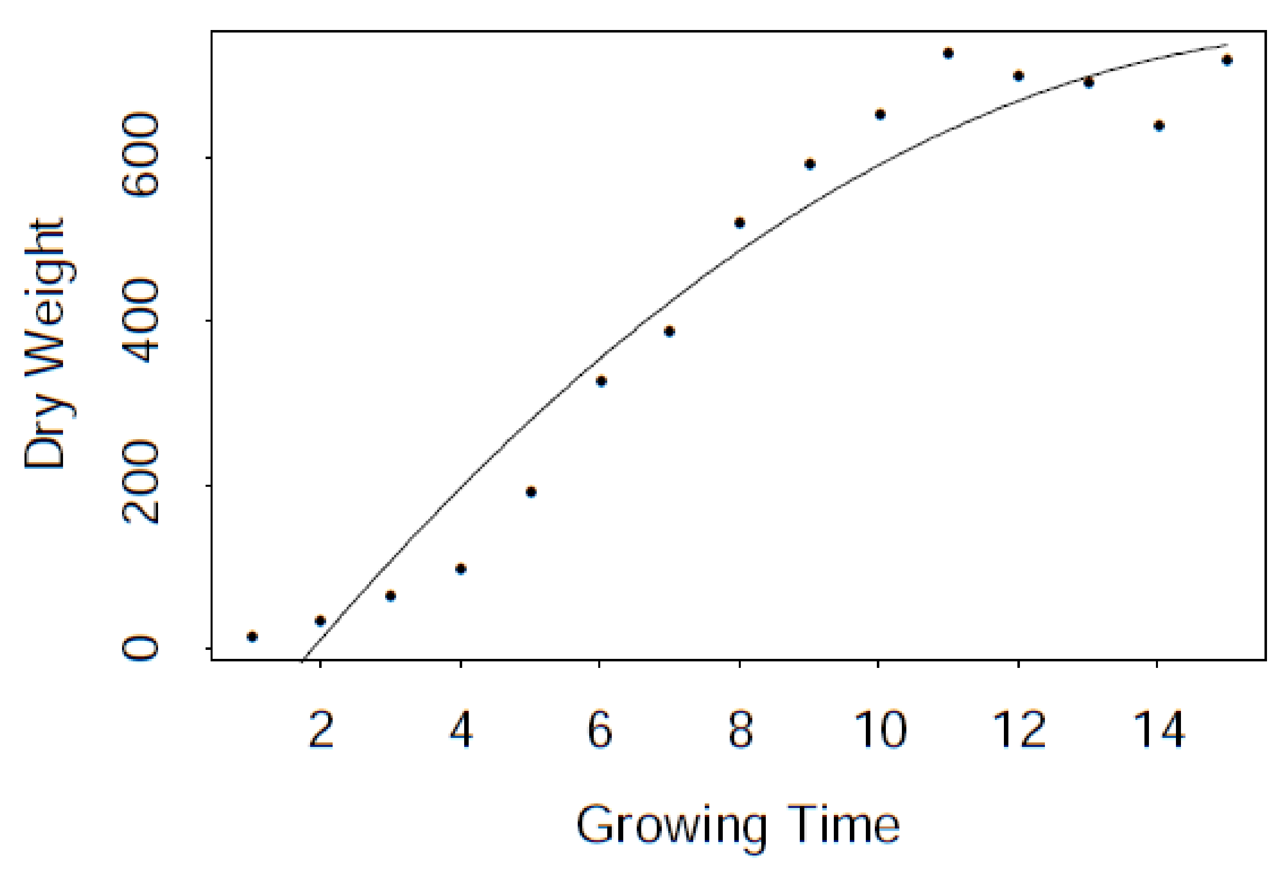
Polynomial Regression

Consider that the following scatterplot display data on the dry weight (Y) of 15 onion bulbs randomly assigned to 15 growing times (X) until measurement.



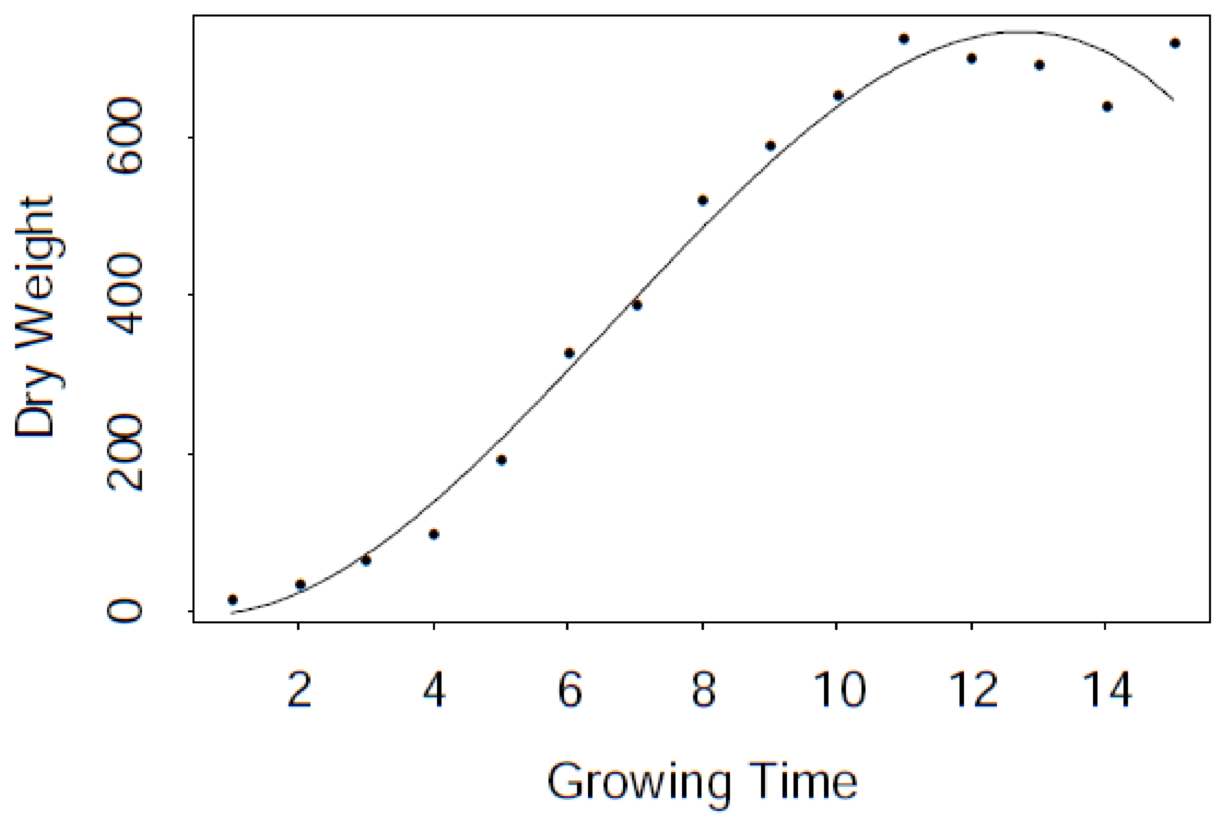
Polynomial Regression

$$Y = \alpha_0 + \alpha_1 x + \alpha_2 x^2$$



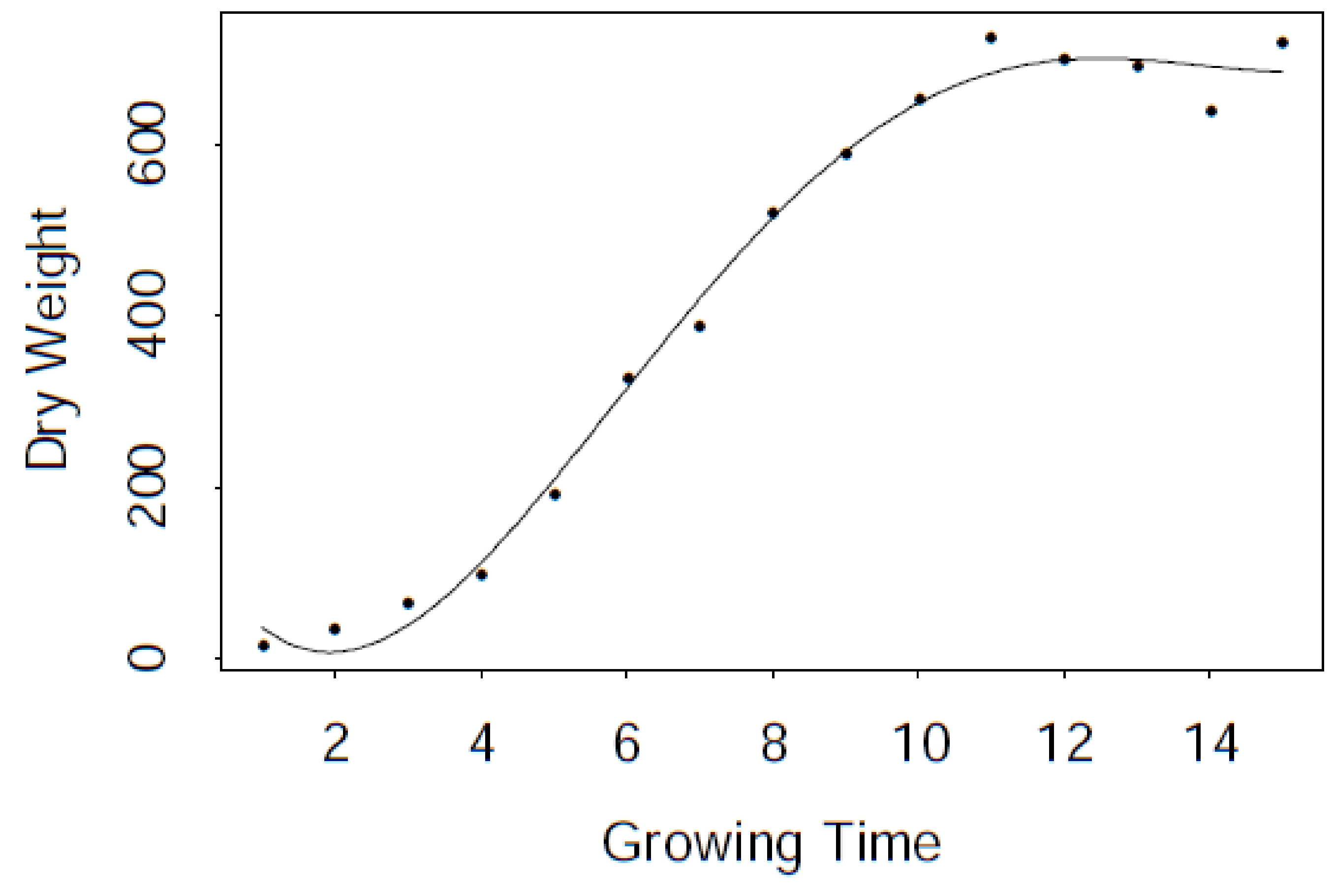
Polynomial Regression

$$Y = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_4 x^4$$



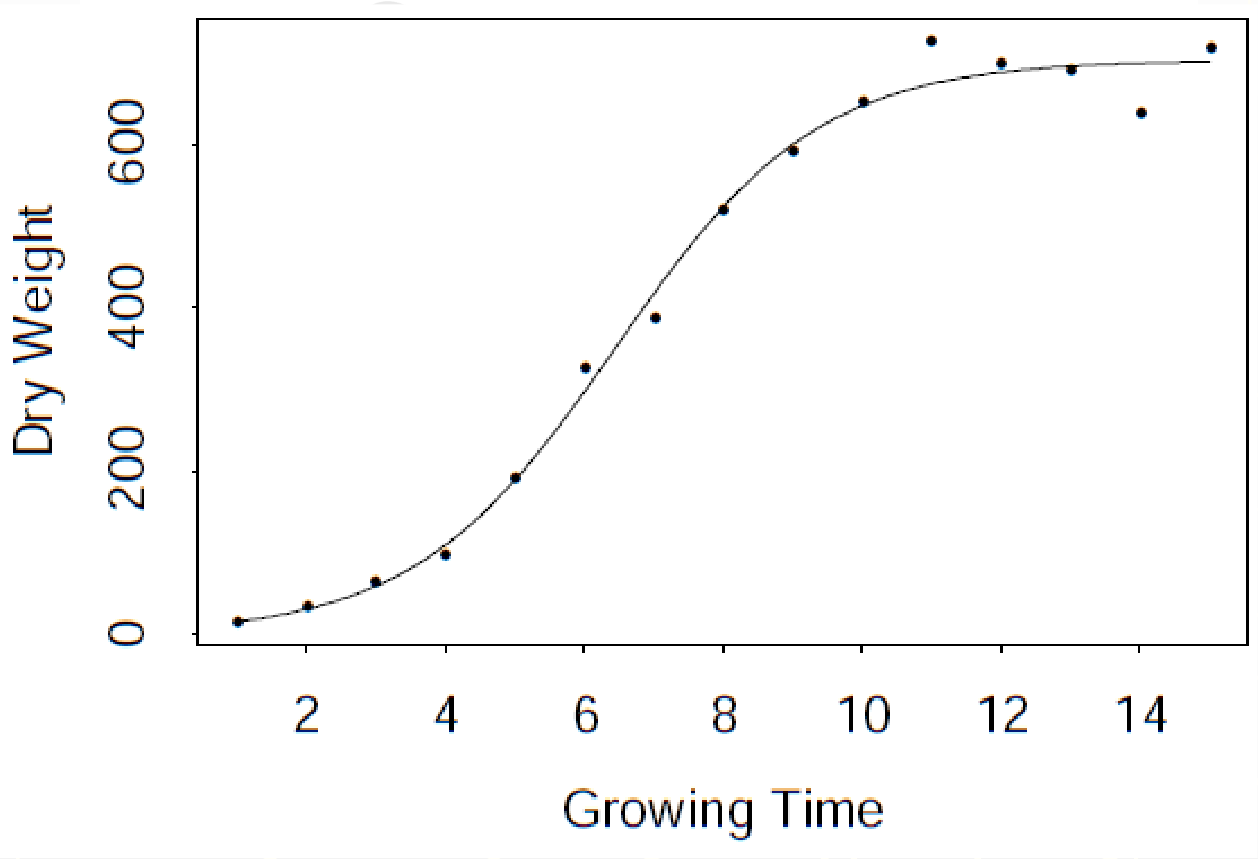
Polynomial Regression

$$Y = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_4 x^4 + \alpha_5 x^5$$

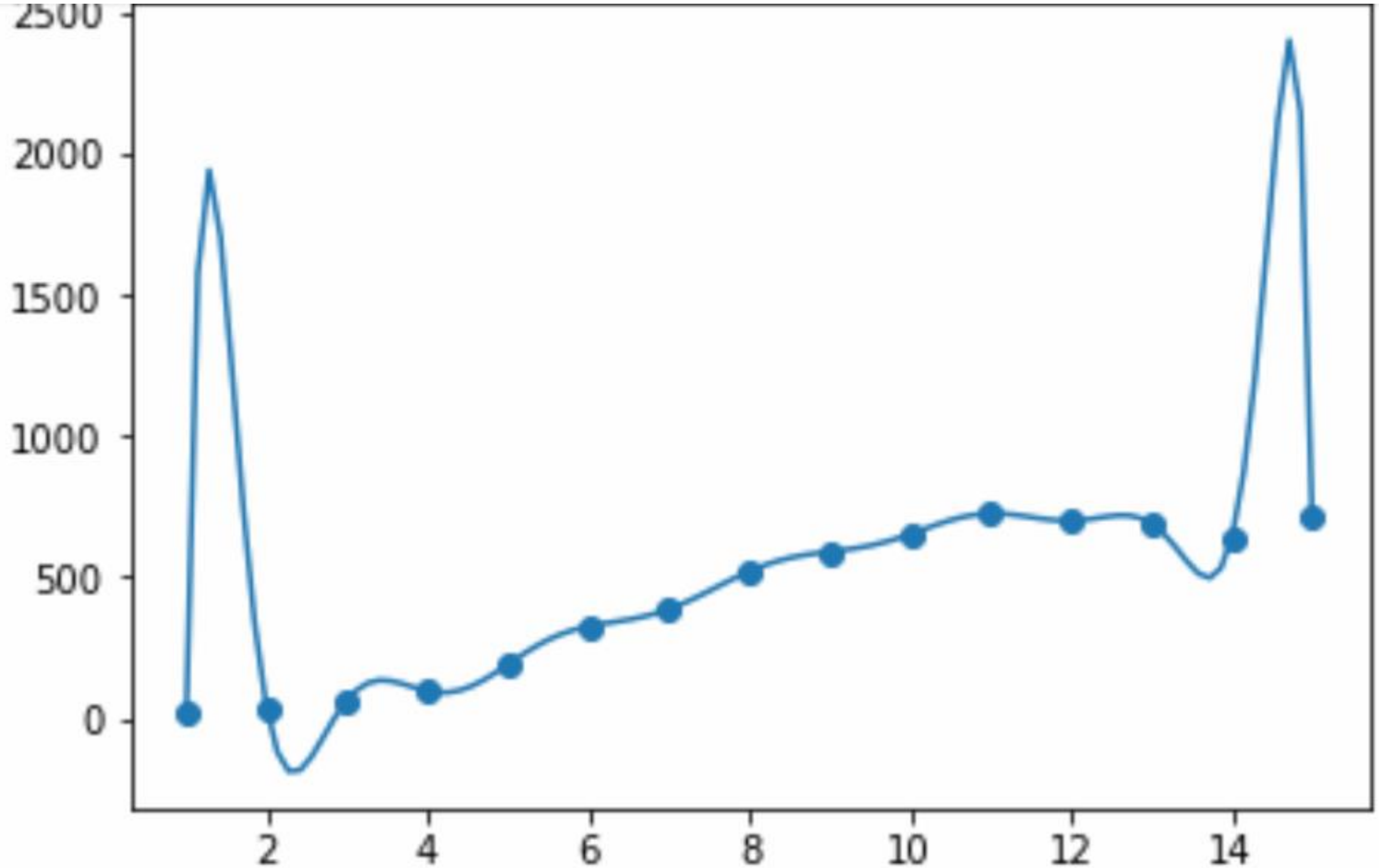


Nonlinear Regression

$$Y = \frac{\alpha_0}{1 + \exp\left\{\frac{\alpha_2 - x}{\alpha_3}\right\}}$$



Overfitting



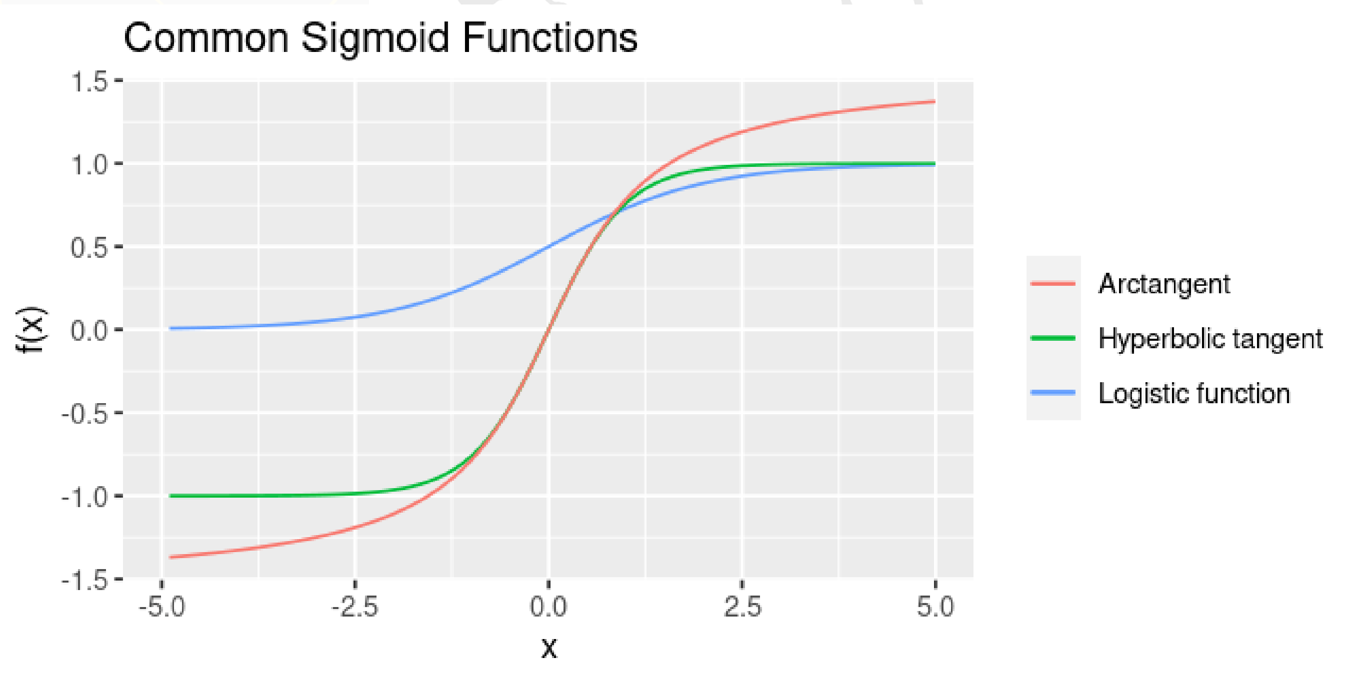
Logistic Regression

Sigmoid Functions

A Sigmoid function is a mathematical function which has a characteristic S-shaped curve.

- logistic function
- hyperbolic tangent
- arctangent

$$S(x) = \frac{1}{1 + e^{-x}}$$



Logistic Regression

Logistic regression is basically a supervised classification algorithm. In a classification problem, the target variable(or output), *y*, can take only discrete values for given set of features(or inputs), *X*.

Contrary to popular belief, logistic regression is a regression model. The model builds a regression model to predict the probability that a given data entry belongs to the category numbered as “1”.

Just like Linear regression assumes that the data follows a linear function, Logistic regression models the data using the sigmoid function.

Logistic Regression-based Classification

Logistic regression becomes a classification technique only when a decision threshold is brought into the picture. The setting of the threshold value is a very important aspect of Logistic regression and is dependent on the classification problem itself.

Based on the number of categories, Logistic regression can be classified as:

binomial: target variable can have only 2 possible types: "0" or "1" which may represent "win" vs "loss", "pass" vs "fail", "dead" vs "alive", etc.

multinomial: target variable can have 3 or more possible types which are not ordered(i.e. types have no quantitative significance) like "disease A" vs "disease B" vs "disease C".

Why Logistic Regression?

Logistic Regression-Maths