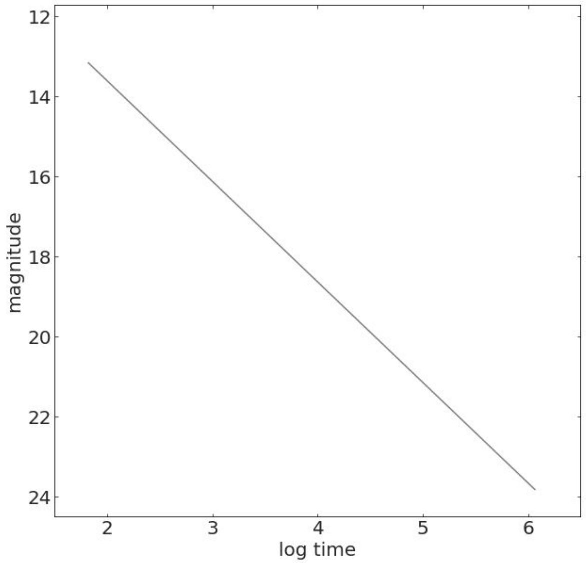
What is Machine Learning?

[Machine Learning is the] field of study that gives computers the ability to learn without being explicitly programmed.

Arthur Samuel, 1959

What is machine learning?



Model:

a mathematical formula with parameters

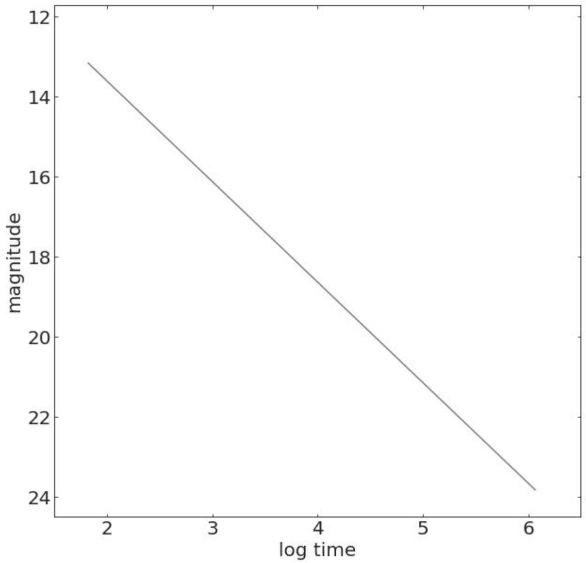
Model

$$y = ax + b$$

parameters: slope (a), intercept (b)

model variable: x - for example time, location, energy

What is machine learning?



Model:

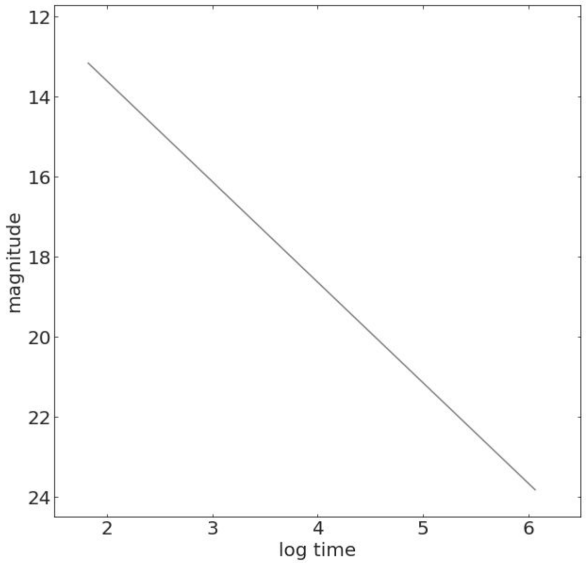
a mathematical formula with parameters

Model

$y = a1x1 + a2x2 + b$

parameters: slope (*a*), intercept (*b*)
model variable: x - for example time, location, energy

What is machine learning?

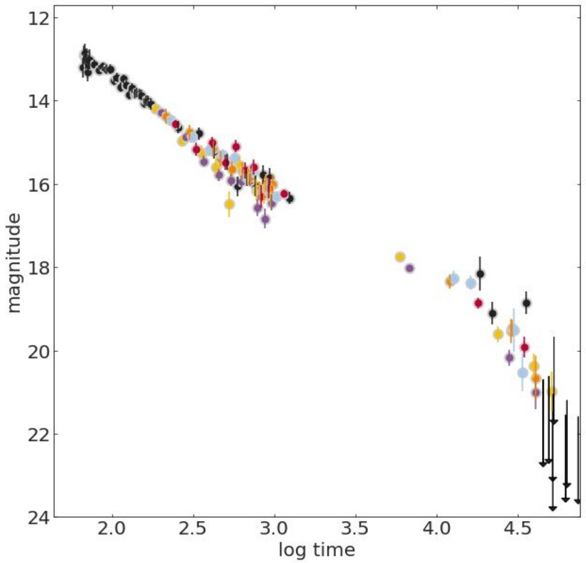


Model:

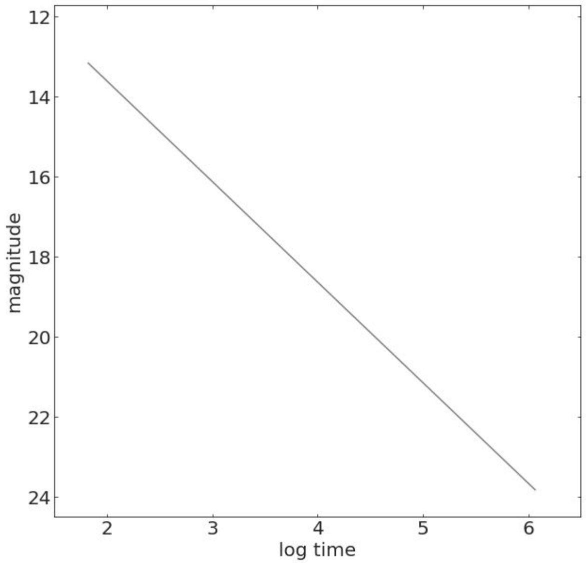
a mathematical formula with parameters

Data:

a set of observations



What is machine learning?

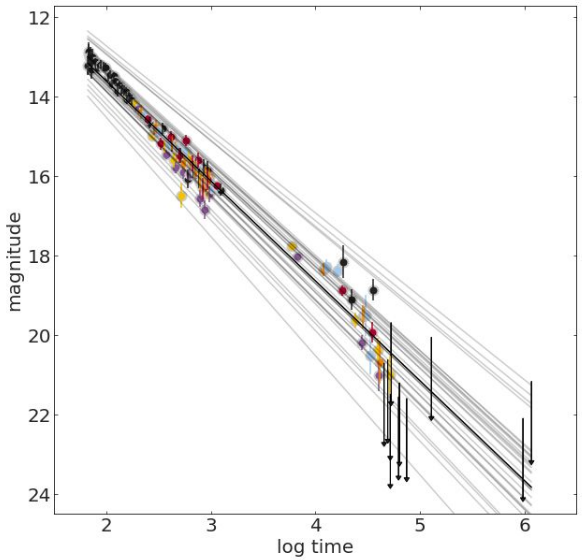


Model:

a mathematical formula with parameters

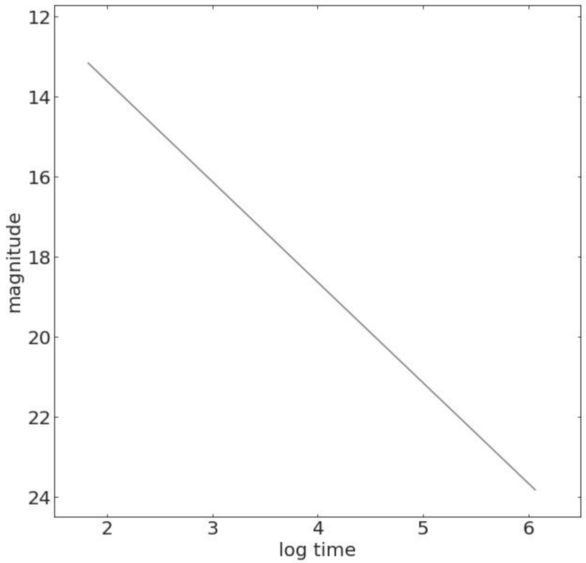
Data:

a set of observations



for every parameter there are an infinity of models

What is machine learning?

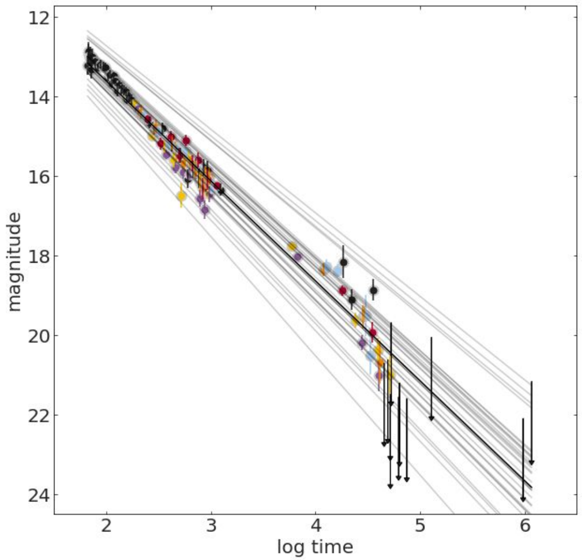


Model:

a mathematical formula with parameters

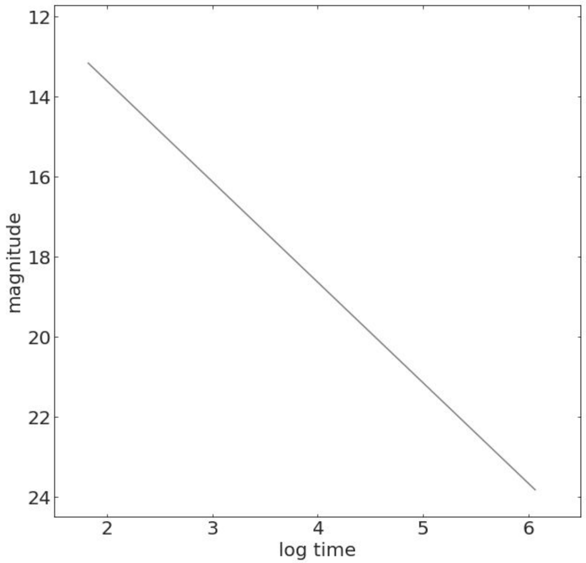
Data:

a set of observations



Use the data to learn the parameters of the model

What is machine learning?

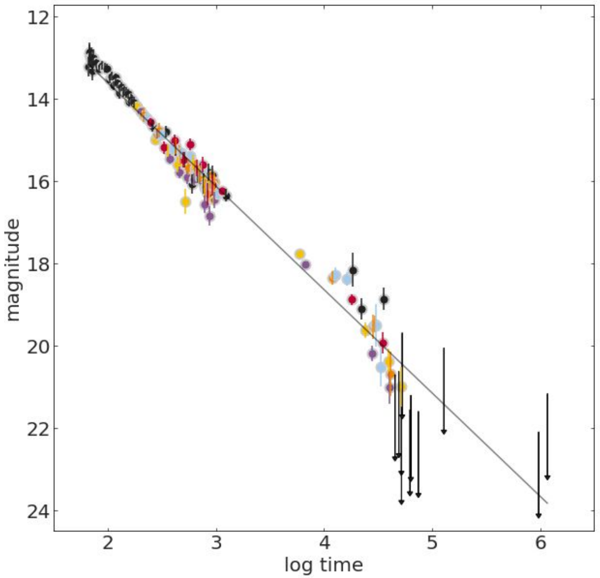


Model:

a mathematical formula with parameters

Data:

a set of observations



Use the data to learn the parameters of the model

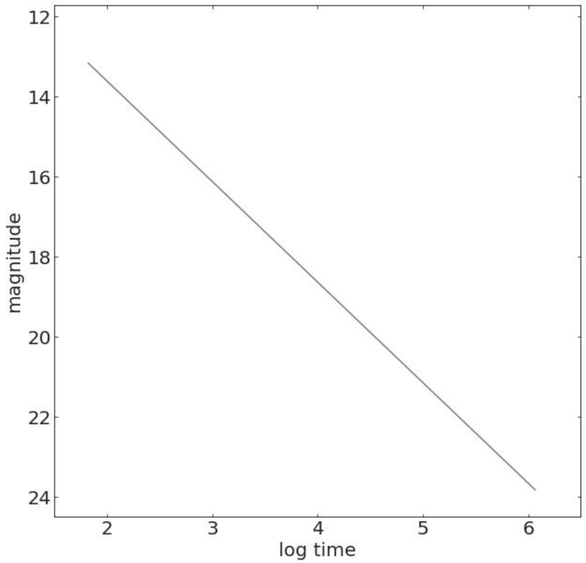
What is machine learning?

Key Concept:

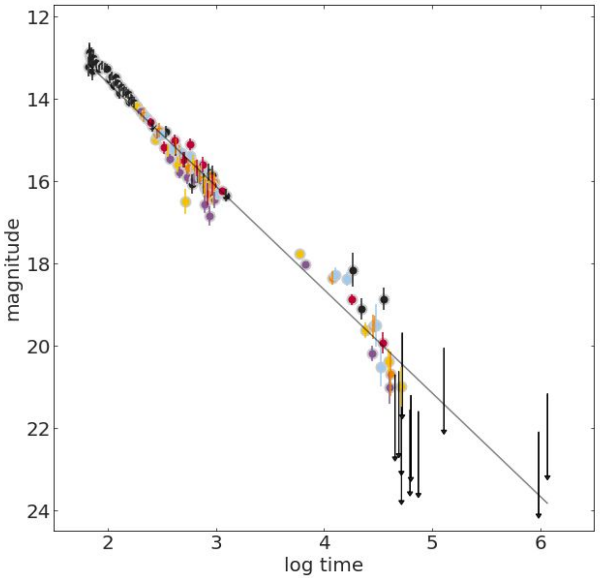
Machine Learning models are parametrized representation of "reality" where the parameters are learned from finite sets of realizations of that reality

Machine Learning is the disciplines that conceptualizes, studies, and applies those models.

What is machine learning?



a model is a low dimensional representation of a higher dimensionality dataset



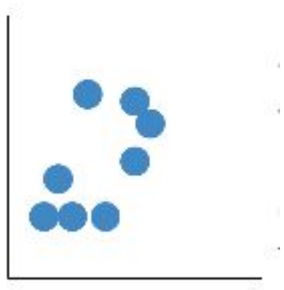
Use the data to learn the parameters of the model

unsupervised vs supervised learning

used to:

- understand structure of feature space
- classify based on examples,
- regression (classification with infinitely small classes)
- understand which features are important in prediction (to get close to causality)

unsupervised vs supervised learning



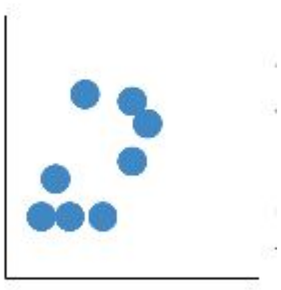
Clustering

partitioning the feature space so that the existing data is grouped (according to some target function!)

unsupervised vs supervised learning

Unsupervised learning

- understanding structure
- anomaly detection
- dimensionality reduction



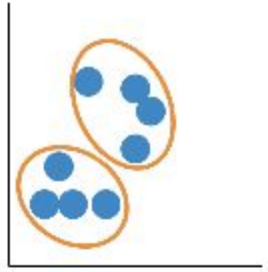
Clustering

partitioning the feature space so that the existing data is grouped (according to some target function!)

unsupervised vs supervised learning

Unsupervised learning

- understanding structure
- anomaly detection
- dimensionality reduction



Clustering

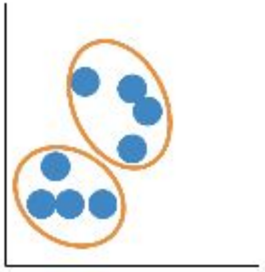
partitioning the feature space so that the existing data is grouped (according to some target function!)

All features are observed for all datapoints

unsupervised vs supervised learning

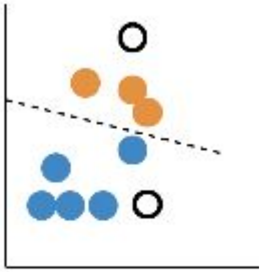
Unsupervised learning

- understanding structure
- anomaly detection
- dimensionality reduction



Clustering

partitioning the feature space so that the existing data is grouped (according to some target function!)



Classifying & regression

finding functions of the variables that allow to predict unobserved properties of new observations.

All features are observed for all datapoints

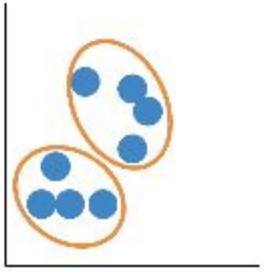
unsupervised vs supervised learning

Unsupervised learning

- understanding structure
- anomaly detection
- dimensionality reduction

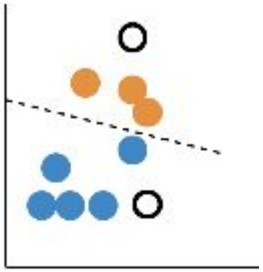
Supervised learning

- classification
- prediction
- feature selection



Clustering

partitioning the feature space so that the existing data is grouped (according to some target function!)



Classifying & regression

finding functions of the variables that allow to predict unobserved properties of new observations.

All features are observed for all datapoints

Some features are not observed for some data points, we want to predict them.

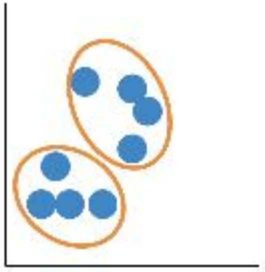
unsupervised vs supervised learning

Unsupervised learning

- understanding structure
- anomaly detection
- dimensionality reduction

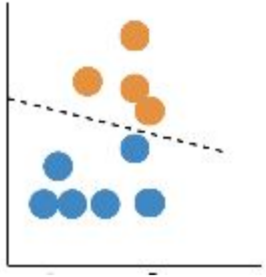
Supervised learning

- classification
- prediction
- feature selection



Clustering

partitioning the feature space so that the existing data is grouped (according to some target function!)



Classifying & regression

finding functions of the variables that allow to predict unobserved properties of new observations.

All features are observed for all datapoints

Some features are not observed for some data points, we want to predict them.

unsupervised vs supervised learning

Unsupervised learning

All features are observed for all datapoints and we are looking for structure in the feature space

Supervised learning

Some features are not observed for some data points: we want to predict them.

The datapoints for which the target feature is observed are said to be "*labeled*"

Semi-supervised learning

A small amount of labeled data is available. Data is cluster and clusters inherit labels

Active learning

The code can interact with the user to update labels.

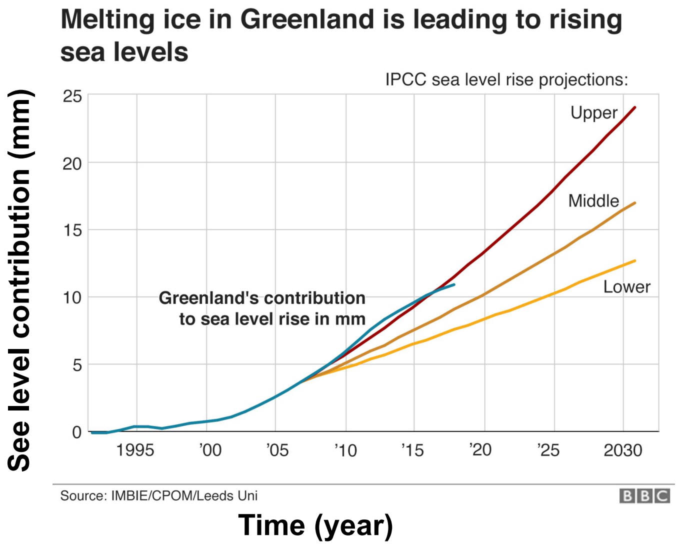
Linear regression

Linear regression

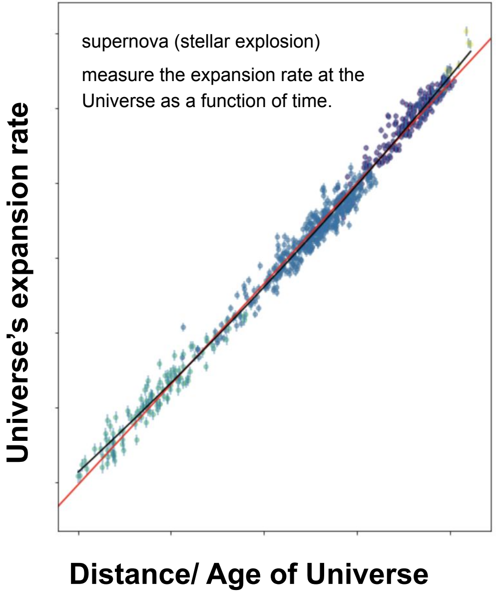
Fitting a line *ax* + *b* to data *y*:

WHY?

To predict and forecast



To explain

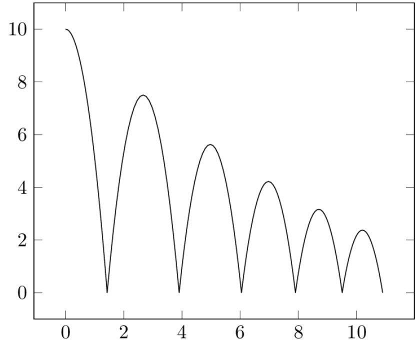


Model Fitting

Predict and forecast: predict the value of the *endogenous* (dependent) variable at locations of the *exogenous* (independent, time) variable where we have no observations. This can be within the observed range, or outside of the range, which in time-series means predict the future (*forecast*)

Explain: relate observed behavior to first principles or behavior of possibly variables to explain the evolution and assess causality.

E.g. fitting a parabola to a bouncing ball demonstrates that gravity (and initial velocity) explains the behavior



Objective function

If there is no analytical solution

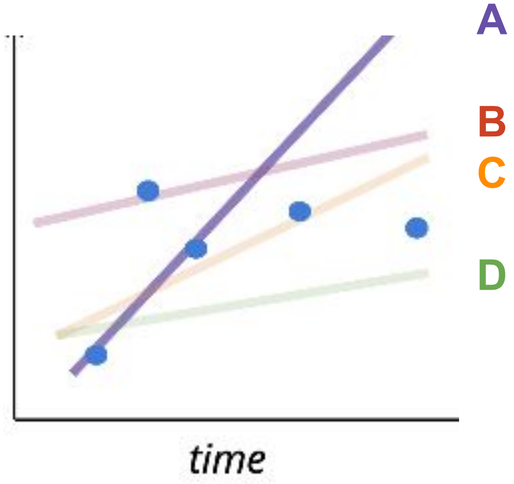
to select the "best" set of parameters we need a plan: we need to choose a function of the parameters to minimize or maximize

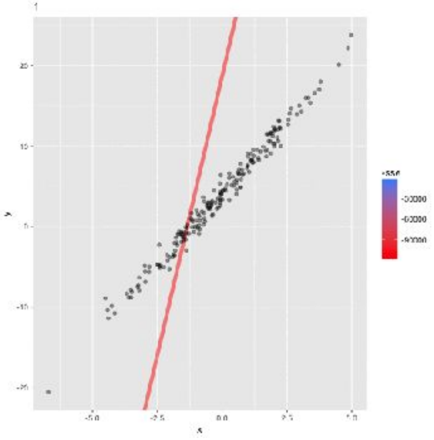
Objective function

If there is no analytical solution

to select the "best" set of parameters we need a plan: we need to choose a function of the parameters to minimize or maximize

Which is the "best fit" line?
A, B, C, D?





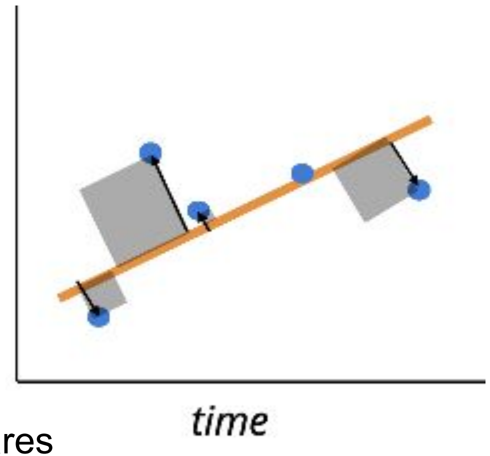
Objective function

If there is no analytical solution

to select the best fit parameters we define a function of the parameters to minimize or maximize → make a generative model for the data

$$L_1 = \sum_{i=1}^{N} |f(x) - y|$$

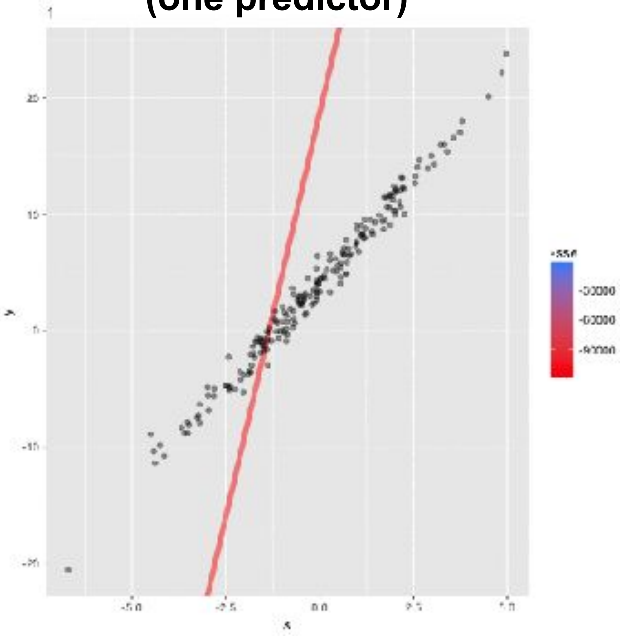
$$L_2 = \sum_{i=1}^{N} (f(x) - y)^2$$



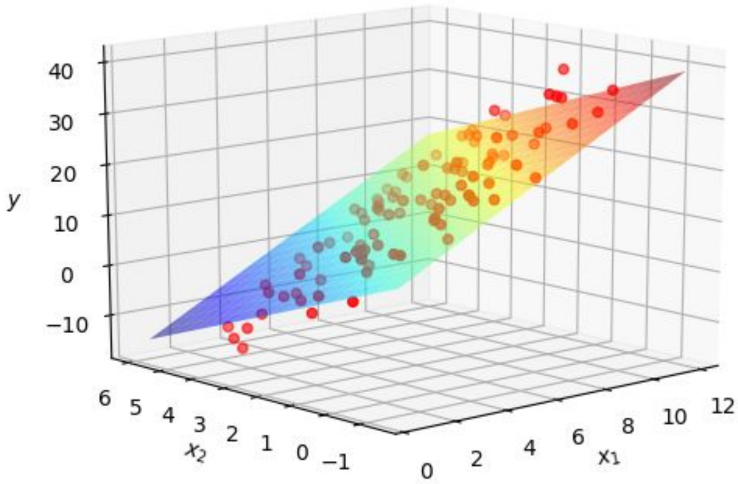
Find the parameters of
f(x) = a* x + b
by minimizing the distance function (L2 = least squares model)

Univariate linear regression vs multivariate linear regression

Univariate
(one predictor)



Multivariate
(multiple predictors)



Multiple linear regression

Many of the models we use in data analysis can be presented using matrix algebra.

We refer to these types of models as *linear models*. “Linear” here does not refer to lines, but rather to linear combinations.

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_r x_r + \varepsilon$$

This equation is the regression equation. $\beta_0, \beta_1, \ldots, \beta_r$ are the regression coefficients, and ε is the random error.

Random component: dependent variable (y) and its probability distribution

Systematic component: explanatory variables (predictor variables)

Link Function: a function of the mean that is a linear function of the explanatory variables

Multiple linear regression

The estimated or predicted response, $f(\mathbf{x}_i)$, for each observation $i = 1, \ldots, n$, should be as close as possible to the corresponding actual response y_i. The differences $y_i - f(\mathbf{x}_i)$ for all observations $i = 1, \ldots, n$, are called the residuals. Regression is about determining the best predicted weights, that is the weights corresponding to the smallest residuals.

To get the best weights, you usually minimize the sum of squared residuals (SSR) for all observations $i = 1, \ldots, n$: SSR = $\Sigma_i(y_i - f(\mathbf{x}_i))^2$. This approach is called the method of ordinary least squares

Multiple linear regression assumptions

- **Weak exogeneity**: the predictor variables *x* can be treated as fixed values, and that they are error-free (do not have measurement errors)

- **Linearity:** the mean of the response variable is a linear combination of the parameters (regression coefficients) and the predictor variables. This does not exclude that the predictor variables can be transformed, used multiple times in different shapes in the link function.

- **Constant variance (or homoscedasticity):** the variance of the errors does not depend on the values of the predictor variables. Thus the variability of the responses for given fixed values of the predictors is the same regardless of how large or small the responses are.

- **Independence of errors**: the errors of the response variables are uncorrelated with each other.

- **Lack of perfect multicollinearity** in the predictors (no linear relationship between two or more predictors)

Multiple linear regression interpretation

Coefficients: Regression coefficients are estimates of the unknown population parameters and describe the relationship between a predictor variable and the response. The sign of each coefficient indicates the direction of the relationship between a predictor variable and the response variable.

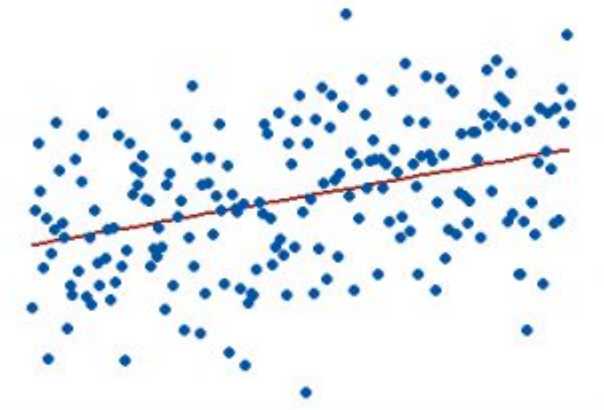
- A positive sign indicates that as the predictor variable increases, the response variable also increases.
- A negative sign indicates that as the predictor variable increases, the response variable decreases.

The coefficient value represents the mean change in the response given a one unit change in the predictor.

Multiple linear regression interpretation

R-squared: A statistical measure that determines the proportion of variance in the dependent variable that can be explained by the independent variable. Simply put, it shows how well the regression model fits the observed data

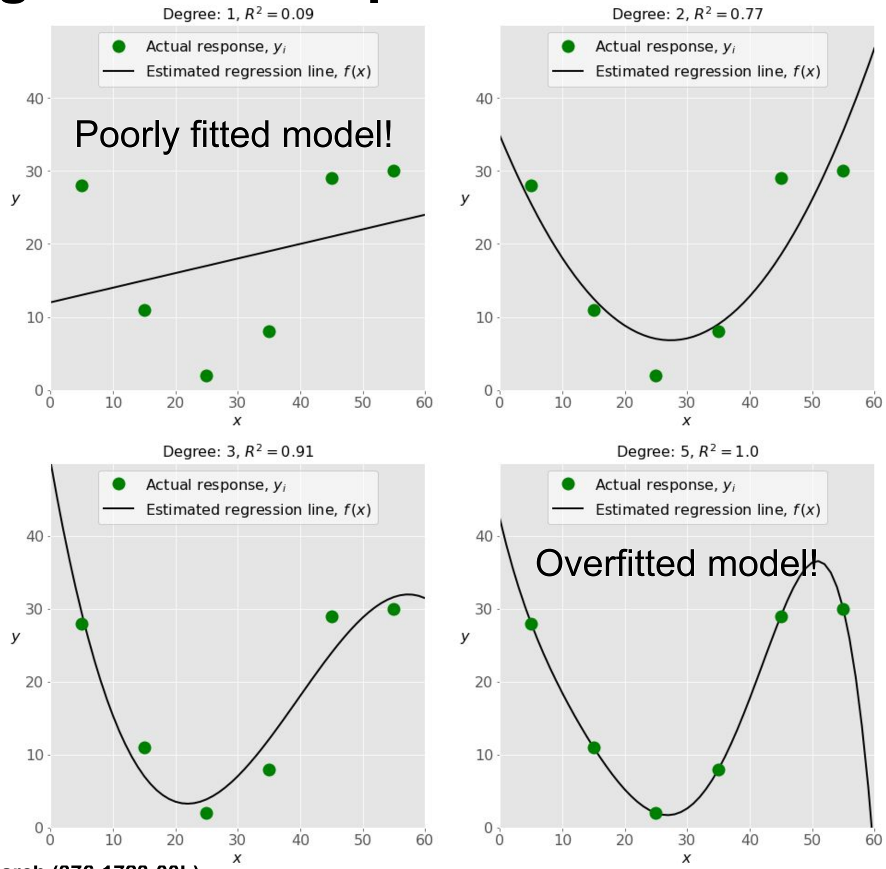
$R^2 = 0.15$



$R^2 = 0.85$

Better predictions!

Multiple linear regression interpretation



Multiple linear regression: example

How is blood glucose affected by other parameters?

How is LDL affected by other parameters?