Supervised ML

Institute of Technology of Cambodia

November 14, 2023

Supervised Learning

Supervised Learning

- Certainly the most successful branch of ML currently
- Training a computer program (algorithm) to learn through examples. The algorithm is trained on labeled data.
- Tasks:
  - Predict the weather, the climate
  - Recognize objects/people in pictures
  - Evaluate the risks of recidivism of a convict (do not do that!)
  - What else?

Supervised Learning

- The goal of supervised learning is to learn a function that can map new input features to the corresponding output labels.
- Two main types of supervised learning
  1. Regression: predict a numerical value
     - Temperature, cost, grade, etc
  2. Classification: predict a class/label/category
     - Success/Failure, Blue/Red/Yellow, which animal among 1000 possible, etc.

Supervised learning: DNN

- Many recent successes thanks to Deep Neural Networks
- This class: only 'classic' methods
- DNN are just an evolution of methods presented in this class, all principles stay the same

Fictional Example

- Let's say we want to predict the happiness of a person. We have a collection of examples, for now in comparable settings (GPD per capita in each country)
- We have access to some characteristics:
  - country, GDP per capita(USD), life satisfaction...
- This is typically a Regression problem

Evaluation/Objective

- Before applying any method, set up an objective/a quality score/an error measure
- We want to be able to compare several prediction methods to see which one is the most efficient. But how to compare them?
- Typical scores:
  - MAE: Mean Absolute Error
  - MSE, RMSE: (Root): Mean Square Error

Mean Squared Error

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - y_i)^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2$$

- Similarity with the **Variance**
- Using squared errors give stronger importance to large errors
  - Strength and weakness (outliers)
- Simple to interpret
  - lower the value, lower the error, better the prediction, can be easier to interpret

$$RMSE = \sqrt{MSE}$$

Mean Absolute Error

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - y_i| = \frac{1}{n}\sum_{i=1}^{n}|e_i|$$

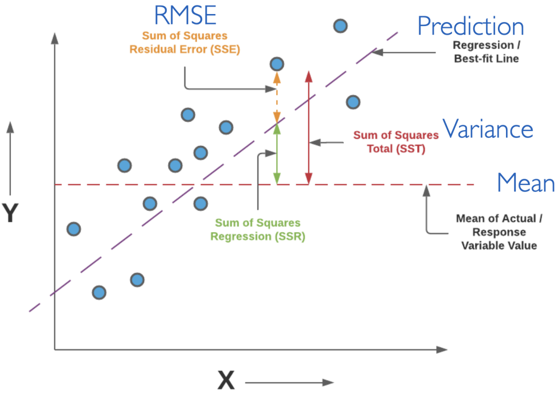
- Similarity with the MAD (Mean Absolute Deviation), comparing values with predictions instead of simple mean
- Simple to interpret
  - lower the value, lower the error, better the prediction
  - 0: perfect prediction

R-Squared

$$R_2 = 1 - \frac{MSE}{Var(y)}$$

- Quantified the fraction of the variance that is explained by the prediction.
- Sometimes called coefficient of determination for linear regression

R-Squared



RMSE
Sum of Squares
Residual Error (SSE)
Prediction
Regression /
Best-fit Line
Sum of Squares
Total (SST)
Variance
Mean
Mean of Actual /
Response
Variable Value
Sum of Squares
Regression (SSR)
Y
X

Evaluation/Objective

Which one should you use?

- Different literature have their favorite one. RMSE is probably the most popular
- If your ML algorithm use the RMSE as objective (loss function), then you should probably use RMSE

If you are not writing a paper or playing a competition, use all of them

- More info can allow you to judge better. There is no 'truth'

Linear Regression

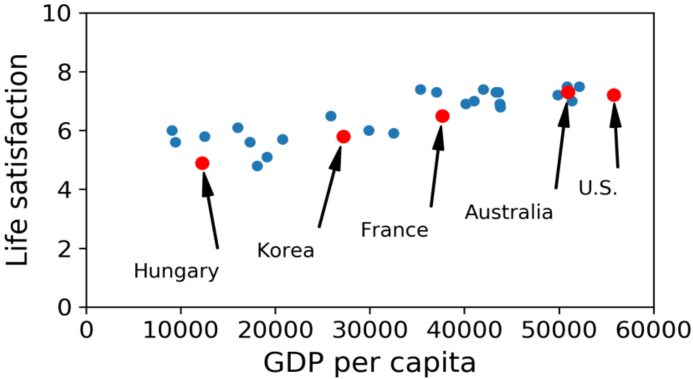
Model the relationship between a **dependent variable** and one or more **independent variables**.

- Dependent variable is the variable that you are trying to predict
- Independent variables are the variable that you think might affect the dependent variable.

This means that a straight line can be used to model the relationship between the variables.

Linear Regression

- Suppose you want to know if money makes people happier?
- There seems to be a linear relationship

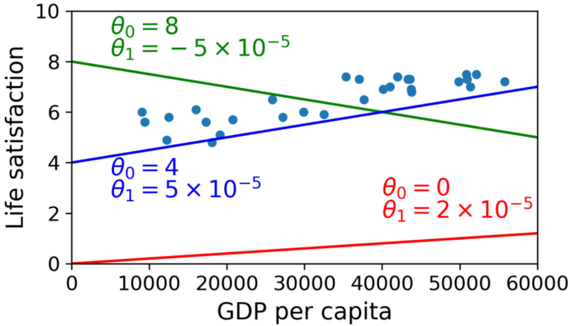


Linear Regression

- It looks like life satisfaction goes up more or less linearly as the country's GDP per capita increases.
- A linear model of life satisfaction with just one attribute, GDP per capita

$$life\_satisfaction = \theta_0 + \theta_1 x GDP\_per\_capita$$

Linear Regression



Linear Regression

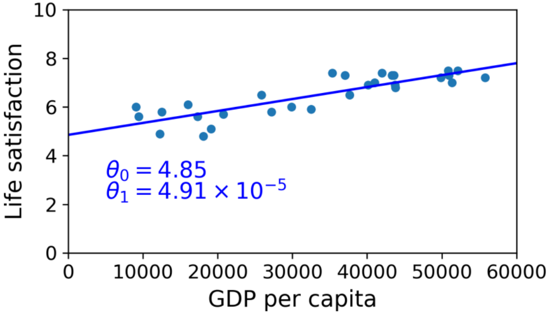
How can you know which values will make your model perform best?

- Define cost function: measure how bad it is
- Define a utility function (or fitness function): measures how good your model is

In linear regression problems, people typically use a cost function that measures the distance between the linear model's predictions and the training examples;the objective is to minimize the distance.

Linear Regression

Now the model fits the training data as closely as possible (for a linear model)



Linear Regression

We will use **linear regression** method, and more specifically **Ordinary Least Square**. First, with a single variable. We assume that

$$y_i = \beta_0 + \beta_1 x_i + \epsilon$$

y is the dependent variable.

β0 is the intercept.

β1 is the slope.

x is the independent variable.

ε is the error term.

- The intercept is the value of y when x is equal to zero
- The slope is the change in y for a one-unit change in x
- The error term is the difference between the observed value of y and the predicted value of y.

Linear Regression

- Linear regression works if there are indeed linear relations
  - But there is no particular reason for relations to be linear
- In many scientific domains (e.g., epidemiology, biology, econometrics, etc.), linear regression is still widely used
  - Why?

OLS Strength

- OLS (Ordinary Least Square Linear Regression) is simple to understand and flexible
  - If I know that a relation between variable and target is log-linear, or other, I can transform my variables beforehand to fall back on a linear problem (but..)
- Linear regression is interpretable:
  - The meaning of each coefficient can be interpreted (positive/negaitve, strength, significance), and thus the relatives strength of the corresponding features
  - True only if strict conditions are respected
    1. No multicolinearity (correlations between variables)
    2. No endogeneity (correlation between variables and errors)
    3. No Homoscedasticity (inhomogeneous distribution of errors)
    4. Etc

OLS?

Common questions: Is OLS machine learning or stat/econometrics?
In my opinion:

- Not a very meaningful questions. Is tomato a vegetable or a fruit?
- If we focus mainly on the coefficients, and not on the prediction. It is not classic machine learning
- If we focus mainly on the prediction and compare with other models. It is a machine learning method like any other