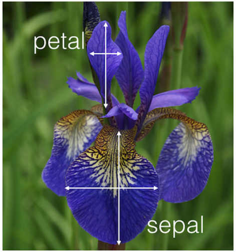
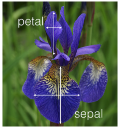
Machine Learning

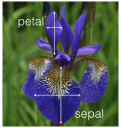
Linear Regression

Alberto Maria Metelli



Credits to Francesco Trovò





Outline

Outline

Practical Information

Alberto Maria Metelli
Postdoctoral Researcher

- Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB)
  - bulding 21
  - first floor
  - office 19
- Email: `albertomaria.metelli@polimi.it`
- Voice: (+39 02 2399) 9679

Practical Information

- Course materials (slides, pdfs, links to recorded lectures) uploaded to WeBeep

Suggested literature to follow the course lectures:

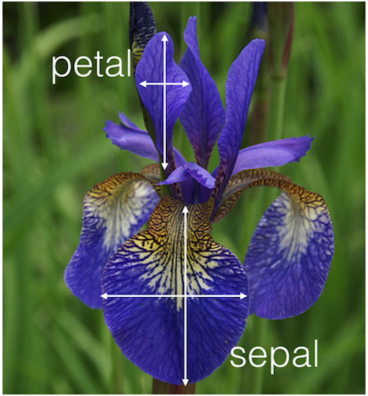
- Bishop, C.M., “Pattern recognition and machine learning”, 2006, Springer
- James, G., Witten, D., Hastie, T., Tibshirani, R., “An introduction to statistical learning”, 2013, Springer

Outline

Where Everything Starts

Consider the **Iris Dataset**:

- Sepal length
- Sepal width
- Petal length
- Petal width
- Species (Iris setosa, Iris virginica e Iris versicolor)



$N = 150$ total samples (50 per species)

Example of Dataset

Sepal length	Sepal width	Petal length	Petal width	Class
5.1000	3.5000	1.4000	0.2000	Iris-setosa
4.9000	3.0000	1.4000	0.2000	Iris-setosa
4.7000	3.2000	1.3000	0.2000	Iris-setosa
4.6000	3.1000	1.5000	0.2000	Iris-setosa
5.0000	3.6000	1.4000	0.2000	Iris-setosa
5.4000	3.9000	1.7000	0.4000	Iris-setosa
4.6000	3.4000	1.4000	0.3000	Iris-setosa
5.0000	3.4000	1.5000	0.2000	Iris-setosa
4.4000	2.9000	1.4000	0.2000	Iris-setosa
4.9000	3.1000	1.5000	0.1000	Iris-setosa
5.4000	3.7000	1.5000	0.2000	Iris-setosa

Scientific Questions

- Can we extract some information from the data?
- What can we infer from them?
- Can we provide predictions on some of the quantities on newly seen data?
- Can we predict the **petal width** of a specific kind of *Iris setosa* by using the **petal length**

Scientific Questions

- Can we extract some information from the data?
- What can we infer from them?
- Can we provide predictions on some of the quantities on newly seen data?
- Can we predict the **petal width** of a specific kind of *Iris setosa* by using the **petal length**

Scientific Questions

- Can we extract some information from the data?
- What can we infer from them?
- Can we provide predictions on some of the quantities on newly seen data?
- Can we predict the **petal width** of a specific kind of *Iris setosa* by using the **petal length**

Solution: Linear Regression

We need to specify:

- Hypothesis space: $\hat{t}_n = f(x_n, w) = w_0 + x_n w_1$
- Loss measure: $J(w, x_n, t_n) = RSS(w) = \sum_n (\hat{t}_n - t_n)^2$
- Optimization method: Least Square (LS) method

where $w \in \mathbb{R}^{M+1}$, $M = 1$

We need to practically apply these operations to the dataset:

- By hand
- With a calculator
- With some code

Solution: Linear Regression

We need to specify:

- Hypothesis space: $\hat{t}_n = f(x_n, w) = w_0 + x_n w_1$
- Loss measure: $J(w, x_n, t_n) = RSS(w) = \sum_n (\hat{t}_n - t_n)^2$
- Optimization method: Least Square (LS) method

where $w \in \mathbb{R}^{M+1}$, $M = 1$

We need to practically apply these operations to the dataset:

- By hand
- With a calculator
- With some code

Outline

Preliminary Operations

- Load Data
- Inspect Data
- Select the interesting Data
- Preprocessing
  - shuffling (`shuffle()`)
  - remove inconsistent data
  - remove outliers
  - normalize or standardize data (`zscore()`)
  - fill missing data

Samples $\{s_1, \ldots, s_N\} \rightarrow$ standardization of sample s:

$$\text{z-score}: \quad s \leftarrow \frac{s - \bar{s}}{S} \qquad \bar{s} = \frac{1}{N} \sum_n s_n \qquad S^2 = \frac{1}{N-1} \sum_n (s_n - \bar{s})^2$$

$$\text{Min-max feature scaling}: \quad s \leftarrow \frac{s - \min_n\{s_n\}}{\max_n\{s_n\} - \min_n\{s_n\}}$$

Linear Regression in Python

- We have many different solutions from different libraries
- In this session we use `sklearn` and `statsmodels`
- Alternatively, we can implement the linear regression by hand ($w^* = (X^\top X)^{-1} X^\top t$)

First option:

- Create a linear model (`LinearRegression()`)
- Fit the model to the data (`fit()`)
- Analyse the results

Evaluating the Results

- Residual Sum of Squares (RSS), Sum Of Squared Errors (SSE):

$$RSS(w) = \sum_n (\hat{t}_n - t_n)^2$$

- Root Mean Square Error: $RMSE = \sqrt{\frac{RSS(w)}{N}}$
- Coefficient of determination: $R^2 = 1 - \frac{RSS(w)}{\sum_n (\bar{t} - t_n)^2}$
- Degrees of Freedom: $dfe = N - M - 1$
- Adjusted coefficient of determination $R^2_{adj} = 1 - (1 - R^2)\frac{N-1}{dfe}$

Mean of the targets: $\bar{t} = \frac{\sum_n t_n}{N}$

Example of `OLS(x, t).fit()` Output

```
Dep. Variable:                 y    R-squared (uncentered):         0.927
Model:                       OLS    Adj. R-squared (uncentered):    0.926
Method:            Least Squares    F-statistic:                    1889.
Date:           Wed, 10 Mar 2021    Prob (F-statistic):          1.56e-86
Time:                   12:28:24    Log-Likelihood:               -16.645
No. Observations:            150    AIC:                            35.29
Df Residuals:                149    BIC:                            38.30
Df Model:                      1
Covariance Type:       nonrobust
=========================================================================
      coef     std err        t       P>|t|       [0.025       0.975]
-------------------------------------------------------------------------
x1    0.9628   0.022       43.467    0.000        0.919        1.007
=========================================================================
Omnibus:                   2.326    Durbin-Watson:                  1.437
Prob(Omnibus):             0.313    Jarque-Bera (JB):               1.852
Skew:                      0.210    Prob(JB):                       0.396
Kurtosis:                  3.347    Cond. No.                        1.00
```

Statistical Tests on Coefficients

Single coefficients:

$$H_0 : w_j = 0 \qquad \text{vs.} \qquad H_1 : w_j \neq 0$$

$$t_{stat} = \frac{\hat{w}_j - w_j}{\hat{\sigma}\sqrt{v_j}} \sim t_{N-M-1}$$

where t_{N-M-1} is the T-Student distribution with $N - M - 1$ degrees of freedom

Overall significance of the model:

$$H_0 : w_1 = \ldots = w_M = 0 \qquad \text{vs.} \qquad H_1 : \exists w_j \neq 0$$

$$F = \frac{dfe}{M} \frac{\sum_n (\bar{t} - t_n)^2 - RSS(w)}{RSS(w)} \sim F_{M,N-M-1}$$

where $F_{M,N-M-1}$ is the Fisher-Snedecor distribution with parameters M and $N - M - 1$

Statistical Tests on Coefficients

Single coefficients:

$$H_0 : w_j = 0 \qquad \text{vs.} \qquad H_1 : w_j \neq 0$$

$$t_{stat} = \frac{\hat{w}_j - w_j}{\hat{\sigma}\sqrt{v_j}} \sim t_{N-M-1}$$

where t_{N-M-1} is the T-Student distribution with $N - M - 1$ degrees of freedom

Overall significance of the model:

$$H_0 : w_1 = \ldots = w_M = 0 \qquad \text{vs.} \qquad H_1 : \exists w_j \neq 0$$

$$F = \frac{dfe}{M} \frac{\sum_n (\bar{t} - t_n)^2 - RSS(w)}{RSS(w)} \sim F_{M,N-M-1}$$

where $F_{M,N-M-1}$ is the Fisher-Snedecor distribution with parameters M and $N - M - 1$

Different Implementations

`fit()`

- Slow in the execution
- Many checks for consistency
- Recap

By-hand solution

- Very fast
- No checks for consistency
- No tests on the coefficients