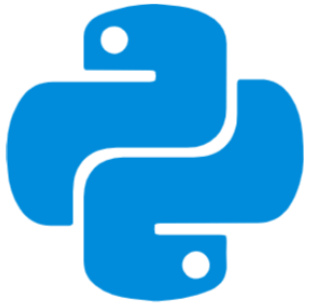
Natural Language Processing
COMP-5413

Lab 1 Prelude

Outline

- **Colab Environment**
  - ✓ Drive mounting and data reading

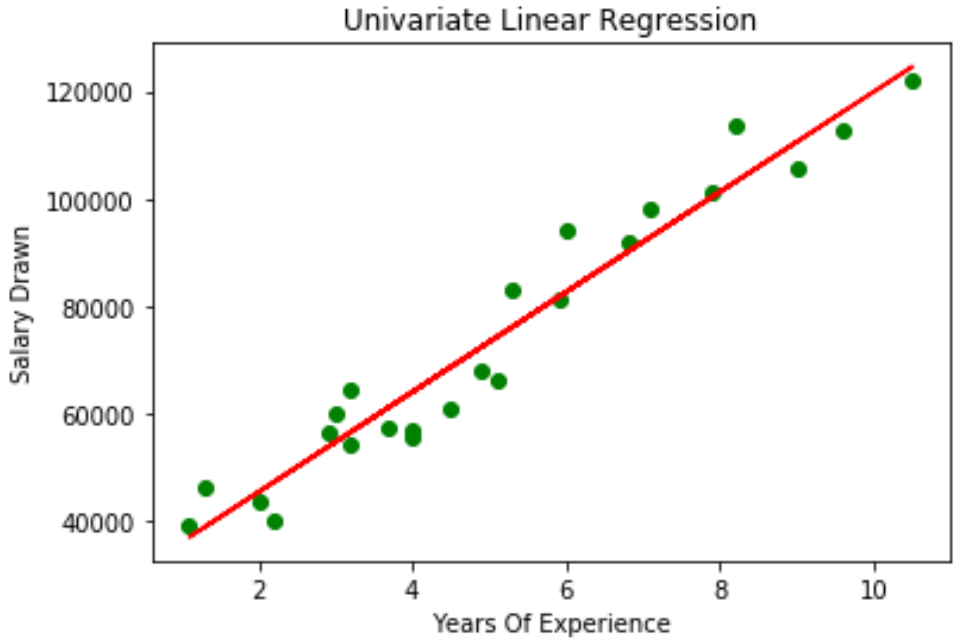
- **Linear Regression Coding Examples**
  - ✓ W/t Scikit-learn libraries, Numpy, and Python
  - ✓ W/t Pytorch



Modeling Sequences – Linear Model

- **The linear predictor function** is a **linear regression model**, where each data point is associated with a continuous outcome $\boldsymbol{y_i}$.
- The general equation for a linear model is: $\boldsymbol{y_i = \sum \beta_i X_i + \epsilon_i}$
- And in **vector notation** is: $\boldsymbol{y_i = \beta^T X_i + \epsilon_i}$,

  - β represents linear parameter estimates to be computed
  - ϵ_i is a disturbance term or error variable.



Modeling Sequences – Linear Model

- And in **vector notation** is: $\boldsymbol{y_i = \beta^T X_i + \epsilon_i}$,

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_n' \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ x_{21} & \cdots & x_{2p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix}, \qquad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

- β represents linear parameter estimates to be computed
- ϵ_i is a disturbance term or error variable.

  - It is a **multivariate input : x11, x12, …, x1p.** It can come from p different sensors.
  - ϵ_i is a disturbance term or error variable — an unobserved random variable that adds noise to the linear relationship between the dependent variable and predictor function.
  - **Benefits** of linear regression: widely used, runs fast, easy to use (not a lot of tuning required), highly interpretable, basis for many other methods

The End

Thank you

Acknowledgment/references

1. https://pytorch.org/get-started/locally/
2. https://colab.research.google.com/notebooks/welcome.ipynb
3. https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html