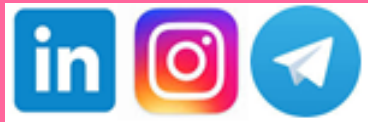


# Machine Learning 7

## Linear Regression



behzad.amanpour





# Linear Regression (Ordinary)

- Regression models the relation between a dependent variable (response) and one or more independent variables (predictor)

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$$

- Each coefficient expresses the impact of a one-unit change in a predictor variable on the mean of the response, provided that all other variables are held constant

<https://statisticsbyjim.com/regression/interpret-coefficients-p-values-regression/>



# Regression vs. Classification

- **Classification** is the task of predicting a discrete class label.
- **Regression** is the task of predicting a quantity.
  - blood pressure, age, ...
- **Logistic regression** models the probability of a class.



# Coefficient of Determination ( $R^2$ )

- **R-squared** value measures how well the model predicts the response.  
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$$
- The best possible score is 1.0, and score 0.0 means constant output.

```
>>> from sklearn.metrics import r2_score
>>> y_true = [3, -0.5, 2, 7]
>>> y_pred = [2.5, 0.0, 2, 8]
>>> r2_score(y_true, y_pred)
0.948...
```

[https://scikit-learn.org/stable/modules/generated/sklearn.metrics.r2\\_score.html](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.r2_score.html)



# Linear vs. Nonlinear Regression

- **Linear**

$$y = b_0 + b_1x_1 + b_2x_2$$

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_1^2 + b_4x_1x_2 + b_5x_2^2.$$

- **Nonlinear**

<https://realpython.com/linear-regression-in-python/>

$$\log y_i = \beta_0 + \beta_1X_{1i} + \beta_2X_{2i} + \varepsilon_i$$

$$y_i = \beta_0 + \beta_1X_{1i} + \frac{1}{\beta_2X_{2i}} + e^{\beta_3X_{1i}X_{2i}} + \varepsilon_i$$

<https://www.mathworks.com/help/stats/what-is-linear-regression.html>