



REGRESSION

CMPE 407

REGRESSION

- Regression searches for relationships among variables.
- For example, you can observe several employees of some company and try to understand how their salaries depend on the **features**, such as experience, level of education, role, city they work in, and so on.
- The dependent features are called the **dependent variables, outputs, or responses**.
- The independent features are called the **independent variables, inputs, or predictors**.
- Why we use is to see how some component influences the other(s)
- <https://realpython.com/linear-regression-in-python/#what-is-regression>

LINEAR REGRESSION

- We use `sklearn.linear_model` import `LinearRegression`
- A linear regression line has an equation of the form $Y = a + bX$, where X is the explanatory variable and Y is the dependent variable. The slope of the line is b , and a is the intercept (the value of y when $x = 0$).
- Linear regression attempts to model the relationship between two (or more) variables by fitting a straight line to the data. Put simply, linear regression attempts to predict the value of one variable, based on the value of another (or multiple other variables).
-

MEAN ABSOLUTE ERROR

- Mean absolute error regression loss.
- Check the other coefficients, etc.
- Prediction Error \rightarrow Actual Value - Predicted Value
- Absolute Error \rightarrow |Prediction Error|

$$mae = \frac{\sum_{i=1}^n abs(y_i - \lambda(x_i))}{n}$$

POISSON REGRESSION

- Poisson regression is used to analyze count data (e.g., the number of drinks per week; the number of arrests per year). Poisson regression is used to answer the questions such as what factors can predict the frequency of an event.
- Compute D^2 , the percentage of deviance explained.
- D^2 is a generalization of the coefficient of determination R^2 . R^2 uses squared error and D^2 deviance. Note that those two are equal for family='normal'.
- D^2 is defined as $D^2 = 1 - D(y_{\text{true}}, y_{\text{pred}}) / D_{\text{null}}$, D_{null} is the null deviance, i.e. the deviance of a model with intercept alone, which corresponds to $y_{\text{pred}} = \bar{y}$. The mean \bar{y} is averaged by sample_weight. Best possible score is 1.0 and it can be negative (because the model can be arbitrarily worse).

REFERENCES

- https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html
- <https://www.displayr.com/what-is-linear-regression/>
- <https://datagy.io/python-sklearn-linear-regression/>
- https://scikit-learn.org/stable/modules/generated/sklearn.metrics.mean_absolute_error.html
- <https://www.sciencedirect.com/topics/psychology/poisson-regression>
- https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.PoissonRegressor.html
- <https://towardsdatascience.com/how-to-build-a-regression-model-in-python-9a10685c7f09>
- https://insidelearningmachines.com/mean_absolute_error/
- https://medium.com/@20__80__/mean-absolute-error-mae-machine-learning-ml-b9b4afc63077