

# Linear Regression Analysis

## CS4372

Roman Hauksson-Neill

Ivan Masyuk

We chose to analyze a dataset called “Combined Cycle Power Plant”, which contains readings of temperature, pressure, relative humidity, and exhaust vacuum of a power plant, which can be used to predict its net hourly electrical energy output.

### Pre-processing

Running the Shapiro-Wilk test on the features, we found that all features are non-normal.

Feature	p-value	Normality
ambient_temp	2.10e-30	Non-normal
vacuum	6.40e-48	Non-normal
ambient_pressure	5.96e-12	Non-normal
relative_humidity	1.47e-28	Non-normal
power_output	6.50e-36	Non-normal

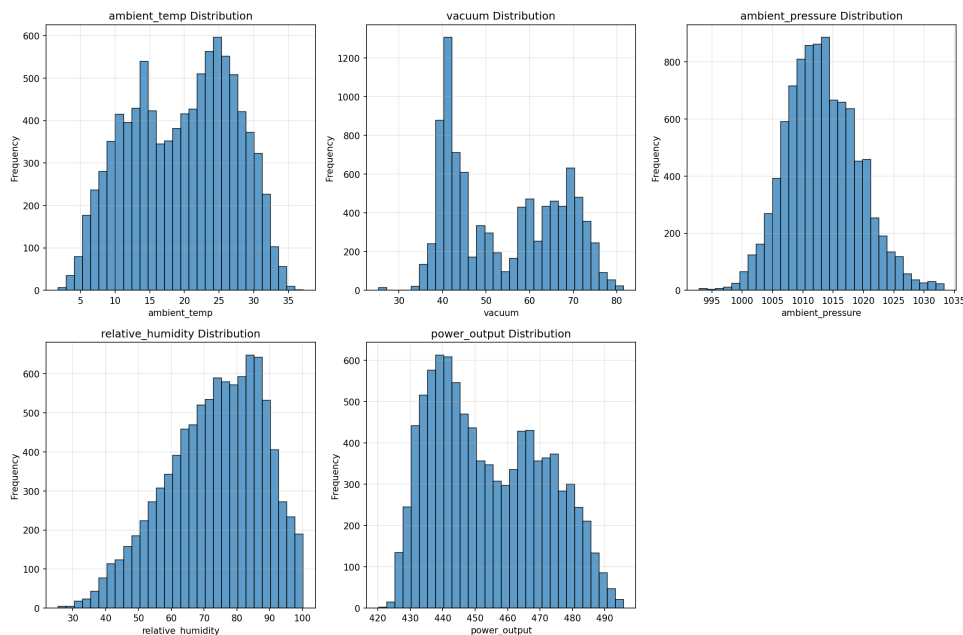


Figure 1: Feature distributions.

Each of the features has a totally different unit and scale. We standardized them so that they all have a mean of 0 and a standard deviation of 1.

Figure 2: Feature distributions after standardization.

### Correlation Analysis

Correlation analysis showed that the target variable (power output) is strongly correlated with the ambient temperature and vacuum features, and the ambient temperature and vacuum features are strongly correlated with each other.

Feature	Correlation	Direction
ambient_temp	-0.948	Negative
vacuum	-0.870	Negative
ambient_pressure	+0.519	Positive
relative_humidity	+0.391	Positive

Figure 3: Correlation matrix.

### Feature selection

Using a simple forward selection approach, we found that the best feature combination is ambient\_temp, vacuum, ambient\_pressure, and relative\_humidity, with a test  $R^2$  of 0.9284.

Features	Test $R^2$	Combination
ambient_temp, vacuum, ambient_pressure, relative_humidity	0.9284	4 features
ambient_temp, vacuum, relative_humidity	0.9282	3 features
ambient_temp, ambient_pressure, relative_humidity	0.9209	3 features
ambient_temp, relative_humidity	0.9209	2 features
ambient_temp, vacuum, ambient_pressure	0.9179	3 features
ambient_temp, vacuum	0.9164	2 features
ambient_temp, ambient_pressure	0.9013	2 features
ambient_temp	0.9000	1 features
vacuum, ambient_pressure, relative_humidity	0.8106	3 features
vacuum, ambient_pressure	0.7960	2 features