Assignment 3. Semiconductor Manufacturer, Logistic Regression (Due-Oct 17th)

Semiconductor manufacturing process dataset

Part I) *We have completed Part I, the data preprocessing, of this project.*

Completed Code (See Example 1 under Practice Problems)*: Data Preprocessing SemiConductor.ipynb (https://canvas.ubc.ca/courses/103654/files/22061133?wrap=1) (https://canvas.ubc.ca/courses/103654/files/22061133/download)*

Part II*), You will build a model to predict the status of a part (Pass/Fail) based on these available data, with 2 different methods.*

- ***Logistic Regression*** *(this is only required for Assignment 3)*
- ***ANN*** *(you will do this for Assignment 4)*

Part III) We will learn how to apply PCA (Principal Component Analyses) for feature selection.

About the Project:

A complex modern semiconductor manufacturing process is normally under constant surveillance via the monitoring of signals/variables collected from sensors and or process measurement points. However, not all of these signals are equally valuable in a specific monitoring system. The measured signals contain a combination of useful information, irrelevant information as well as noise. Engineers typically have a much larger number of signals than are actually required. If we consider each type of signal as a feature, then feature selection may be applied to identify the most relevant signals. The Process Engineers may then use these signals to determine key factors contributing to yield excursions downstream in the process. This will enable an increase in process throughput, decreased time to learning, and reduce per-unit production costs. These signals can be used as features to predict the yield type. And by analyzing and trying out different combinations of features, essential signals that are impacting the yield type can be identified.

Dataset*:* **SemiconductorManufacturingProcessDataset.csv (https://canvas.ubc.ca/courses/103654/files/22061201?wrap=1) (https://canvas.ubc.ca/courses/103654/files/22061201/download)**

Data Source: *https://www.kaggle.com/saurabhbagchi/fmst-semiconductor-manufacturing-project* ⬀
(https://www.kaggle.com/saurabhbagchi/fmst-semiconductor-manufacturing-project)

Assignment 3. Build a Logistic Regression Model. Calculate the Performance, using Confusion Matrix, for your test_set (20% of the data).

Submit Assignment 3. Logistic Regression
(https://canvas.ubc.ca/courses/103654/assignments/1325597)

The deadline is Oct 17th (Monday) at 11:59 pm.

Solution: