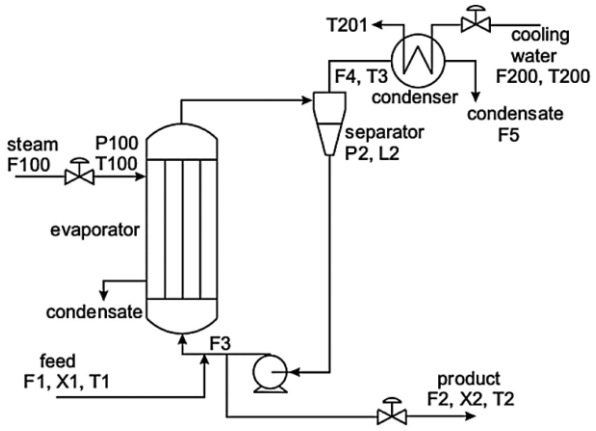
ChE 197/297: Intro to AI/ML for Chemical Engineers
Case Studies in ChemE

Instructions: Answer each problem then create a solution using Python code via Jupyter Notebook.

Problem: Fault Classification in an Evaporator System



Variable	Description	Value	Unit
F_1	Feed flowrate	10.0	kg/min
F_2	Product flowrate	2.0	kg/min
F_3	Circulating flowrate	50.0	kg/min
F_4	Vapor flowrate	8.0	kg/min
F_5	Condensate flowrate	8.0	kg/min
X_1	Feed composition	5.0	%
X_2	Product composition	25.0	%
T_1	Feed temperature	40.0	°C
T_2	Product temperature	84.6	°C
T_3	Vapor temperature	80.6	°C
L_2	Separator level	1.0	m
P_2	Operating pressure	50.5	kPa
F_{100}	Steam flowrate	9.27	kg/min
T_{100}	Steam temperature	119.9	°C
P_{100}	Steam pressure	194.7	kPa
Q_{100}	Heat duty	339.2	kW
F_{200}	Cooling water flowrate	208.0	kg/min
T_{200}	Inlet CW temperature	25.0	°C
T_{201}	Outlet CW temperature	46.15	°C
Q_{200}	Condenser duty	308.0	kW

Consider the evaporator system above. The aim of the system is to take the feed at flow rate F1 and solute concentration X1, then concentrate it to X2 flowing at a rate F2. The vapor is split in a separator: one stream is cooled into a condensate stream and the other is forced back into the evaporator using a pump. Heating is supplied by steam entering at F100 flow rate, T100 temperature and pressure P100.

Our task is fault classification. Given a data set consisting of normal-operation samples and pump-failure samples (**evap_data.csv**), do the following:

1. Split the data into a training (60%), validation (20%), and test set (20%).

2. Normalize the training data, then apply the same scaling to the validation and test data.

3. Train an SVM multi-class classifier on the training set. Use the validation set to find the best values of the box constraint C and kernel length-scale (γ), i.e. find the settings that gives the best validation accuracy. After optimizing the SVM, evaluate the performance of the optimal model on the test set.

END OF EXERCISE