JM0100: Business Analytics
Assignment 3: Group project
2019-2020

Instructions

- Use Python packages to solve the tasks.
- Deadline for submitting the solutions is Thursday, 30 April, 18:00.

Deliverables

Submit all solutions electronically through Canvas. Make sure that you submit everything in a single zip file that contains the solutions to the project, described below:

- Submit a pdf file D-XXX.pdf (where XXX is your group number) that contains your answers to the questions.
- In addition, submit the Python or Jupyter notebook file (D-XXX.py or D-XXX.ipynb) that you created for the project.
- Obey the page limit! You can write *at most 3 pages* for the prediction task, *at most 2 pages* for the first optimization task, and *at most 1 page* for each of the remaining tasks.
- Make sure you include a cover page with the team information, i.e., names and student id of the team members.

Name your zip file D-XXX.zip where XXX is your group number.

Good luck!

Introduction

In this assignment we will consider a maintenance scheduling problem for airplane engines. Engines are subject to wear and tear over time. A key quantity of interest for machines like engines is called the remaining useful lifetime or RUL for short. In this assignment you will develop a predictive model that predicts the RUL for engines that are currently in use. Using an engine beyond its RUL is not safe and therefore there are costs associated with this. The only way to ensure that an engine is safe for use, is to perform maintenance on it. Given an estimate of the RUL, company X needs to decide on a maintenance schedule so that it can maintain the engines in a timely fashion in order to avoid the aforementioned costs.

You are provided with the following files:

1. `DataTrain.txt`. This is a txt-file that contains data for 100 engines. You will use this data to develop your predictive model. We will refer to this data as *dataset 1*. This dataset contains run-to-failure data for a number of engines.

   Engines in *dataset 1* are considered to start with various degrees of initial wear but are considered healthy at the start of each record. As the number of cycles increases the engines begin to deteriorate until they can no longer function. At this point in time (when the engines can no longer function) the engines are considered unhealthy and cannot perform their intended function. A description of the variables in this dataset is given in Table 1.

   Table 1: Dataset description for *dataset 1*.

Column index	Data field	Type	Description
1	engine id	Integer	aircraft engine identifier, range [1, 100]
2	cycle	Integer	time, in cycles
3	setting1	Double	operational setting 1
4	setting2	Double	operational setting 2
5	setting3	Double	operational setting 3
6	s1	Double	sensor measurement 1
7	s2	Double	sensor measurement 2
...	...		
26	s21	Double	sensor measurement 21

   The interpretation of the data rows are as follows. For example, the engine with engine id 1, was working for 192 cycles and then it failed. So, (i) at cycle 191 the RUL for engine with engine id 1 was equal to 1; (ii) at cycle 190 the RUL for engine with engine id 1 was equal to 2.

2. `DataSchedule.txt`. This is a txt-file that contains data for 100 engines that are currently in use. We will refer to this data as *dataset 2*. Unlike *dataset 1*, *dataset 2* contains temporal data that terminates some time before a system failure. This dataset contains the same variables as *dataset 1*. The RUL for these engines is not known. Based on your predictive model, you will predict the RUL for these 100 engines.

3. `RUL_consultancy_predictions_A3.csv`. The contents of this file will be explained in Optimization Task 1.

Prediction Task

In this task the goal is to develop a predictive model that will predict the remaining useful lifetime RUL of the engines.

Part A

1. Use *data set 1* to develop a model that can predict the RUL of an engine.
2. Clearly describe the steps in your approach. Some aspects that you can think about are as follows. How do you split your dataset for training and testing? Which features do you construct? How do you assess the performance of your model? Which types of models do you consider and why?

Part B

In this task you are going to use your prediction model to predict the RUL for engines that are currently in use.

1. Use your prediction model from Part A and predict the RUL of the engines in *data set 2*.
2. Please round your predictions to integer values. These integer valued predictions will be used in the subsequent tasks.

Optimization Task 1

Introduction

In this task, the goal is to develop a maintenance schedule for the engines in *data set 2*. The output of the previous task should be a list that contains a predicted RUL for each engine in *data set 2*. These predicted values will subsequently be used as input in order to allocate teams of workers to perform maintenance tasks.

Maintenance teams

We want to allocate workers to different engines in order to perform maintenance. Workers perform maintenance in teams and teams can only work on one engine at a time. Also, two teams cannot work on the same engine simultaneously. There are different types of teams: teams of type A and teams of type B. Teams of type A are more efficient than teams of type B in the sense that it takes them μ_j^A days to perform maintenance on engine $j \in \mathcal{M}$ and $\mu_j^A < \mu_j^B$. Here $\mathcal{M}$ is the set of indices of the engines on which maintenance needs to be performed. There are G teams in total and assume that every team can be associated with a unique number from the set $\mathcal{G} = \{1, \dots, G\}$. Let the set indices of the teams of type A be denoted by $\mathcal{G}^A$. Let the set indices of the teams of type B be denoted by $\mathcal{G}^B$. Then the indices of all the teams is given by $\mathcal{G} = \mathcal{G}^A \cup \mathcal{G}^B$ with $\mathcal{G}^A \cap \mathcal{G}^B = \varnothing$.

Timing of events, penalty costs and safety due date

Time is discrete and measured in days. For each engine we associate a *safety due date* with its RUL. Suppose that we are currently at day $t = t_1$ and that engine j has a RUL of R_j, then the safety due date is equal to day $t = t_1 + R_j - 1$, that is, on day $t = t_1 + R_j - 1$ the engine is still working properly and on day $t > t_1 + R_j - 1$ the engine is past its safety due to date and not working properly.

If an engine is not maintained by its safety due date, then the company incurs a penalty cost. The cost depends on the amount of days since the safety due date: for each day after the safety due date, the company incurs a cost of c_j for engine j.

In this task a maintenance schedule is made for a planning horizon of T days. That is, if we are currently at day $t = t_1$ and the planning horizon is T, then the maintenance schedule pertains to the planning period which consists of days $t_1, t_1 + 1, \ldots, t_1 + T - 1$.

In this task we make the following assumption: if maintenance is performed on an engine during the planning period, then the RUL (after maintenance) of the engine will exceed T. This ensures that engines cannot "fail" twice within a planning horizon (that is, engines cannot have two safety due dates within a planning horizon). As a consequence we have the following: maintenance on engine j can be performed at most once during the planning period.

Since working with the exact values of the RUL may be to complex, the company has developed a classification system for the engines based on the RUL. Engines with $R_j \leq 25$ have label "danger", engines with $R_j \leq 40$ have label "caution", the others have label "Normal".

The company is only interested in *complete* schedules: if a team is assigned to perform maintenance during the planning period, then the team must be able to complete the maintenance within the planning period.

Example 1. *In this example we illustrate the concepts discussed above in order to avoid confusion. Suppose that:*

1. *The planning horizon is* $T = 20$.
2. *We are at day* $t = 1$.
3. *Engine* 1 *has a RUL of* 5*, engine* 2 *has a RUL of* 7 *and engine* 3 *has a RUL of* 18.
4. *It takes* 6 *days for a team of type A to perform maintenance on engine* 1*, that is* $\mu_1^A = 6$.
5. *It takes* 4 *days for a team of type A to perform maintenance on engine* 2*, that is,* $\mu_2^A = 4$.
6. *It takes* 4 *days for a team of type A to perform maintenance on engine* 3*, that is,* $\mu_3^A = 4$.
7. *Team* A_1 *of type A starts performing maintenance on engine* 2 *on day* $t = 9$.
8. *Team* A_2 *of type A starts performing maintenance on engine* 3 *on day* $t = 18$.
9. *Team* A_3 *of type A starts performing maintenance on engine* 1 *on day* $t = 3$.

Under these specifications, we have the following:

1. $\mathcal{M} = \{1, 2, 3\}$. *The planning period consists of days* $1, 2, 3, 4, \ldots, 20$.
2. *The safety due date for engine* 1 *is* $t = 5$*: engine* 1 *is working properly on days* $1, 2, 3, 4, 5$ *but not on day* $t \geq 6$.
3. *On day* $t = 8$ *engine 2 is 1 day past its safety due date. The penalty costs (during the planning period) for engine 2 are* $1 \cdot c_2$.
4. *On day* $t = 9$ *engine 2 is 2 days past its safety due date. The penalty costs (during the planning period) for engine 2 are* $2 \cdot c_2$.
5. *On days* $9, 10, 11, 12$ *maintenance is performed on engine* 2 *by team* A_1.
6. *On days* $9, 10, 11, 12$ *team* A_1 *can only work on engine* 2.
7. *On days* $9, 10, 11, 12$ *only team* A_1 *is allowed to work on engine* 2.

8. *On day $t = 12$ engine 2 is 5 days past its safety due date. The penalty costs (during the planning period) for engine 2 are $5 \cdot c_2$.*

9. *On day $t = 13$ engine 2 is working properly. The penalty costs (during the planning period) for engine 2 are $5 \cdot c_2$. On day $13 < t \leq 20$ engine 2 is working properly. The penalty costs (during the planning period) for engine 2 are $5 \cdot c_2$.*

10. *On days $18, 19, 20, 21$ maintenance is performed on engine 3 by team A_2. This is not allowed in a complete schedule because maintenance will be completed on day 21 and this is not in the planning period (which consists of days $1, 2, 3, 4, \ldots, 20$).*

11. *Suppose that we have the following schedule:*

    (a) *Team A_1 of type A starts performing maintenance on engine 2 on day $t = 9$.*

    (b) *Team A_3 of type A starts performing maintenance on engine 1 on day $t = 3$.*

    (c) *No team is allocated to engine 3.*

    Then, the penalty costs for engine 1 equals $3 \cdot c_1$, the penalty costs for engine 2 equals $5 \cdot c_2$, the penalty costs for engine 3 equals $2 \cdot c_3$. So the total costs for this schedule is: $3 \cdot c_1 + 5 \cdot c_2 + 2 \cdot c_3$.

$\square$

Part A

In this task company X is going to allocate teams to different engines in order to perform maintenance. Assume that we are currently at day $t = 1$ and that company X wants to allocate teams to engines in order to minimize penalty costs for a planning horizon of $T = t_p$. More specifically, company X wants to allocate teams to engines that have a predicted RUL that satisfies $R_j \leq T$.

1. Company X wants to make a schedule that allocates teams to different engines in order to perform maintenance during the planning horizon. Formulate this as an Mixed Integer Linear Programming (MIP) problem.

2. Use clear and correct notation. Clearly define (i) the decision variables, (ii) objective function(s), (iii) constraint(s), (iv) parameters of the problem, etc.

Part B

In this task you are going to solve the optimization problem that you have formulated in the previous part.

Assume that there are 4 teams: 2 teams of type A and 2 teams of type B. Assume that $\mathcal{G}^A = \{1, 2\}$, $\mathcal{G}^B = \{3, 4\}$, $\mathcal{G} = \{1, 2, 3, 4\}$.

The values for $c_j, j \in \mathcal{M}$ are as follows: $c_j = 5, j \in \{1, 2, \ldots, 20\}$, (ii) $c_j = 7, j \in \{21, 22, \ldots, 40\}$, (iii) $c_j = 9, j \in \{41, 42, \ldots, 60\}$, (iv) $c_j = 5, j \in \{61, 62, \ldots, 80\}$ and (v) $c_j = 3, j \in \{81, 82, \ldots, 100\}$.

The values for $\mu_j^A, j \in \mathcal{M}$ are as follows: $\mu_j^A = 4, j \in \{1, 2, \ldots, 25\}$, (ii) $\mu_j^A = 6, j \in \{26, 27, \ldots, 50\}$, (iii) $\mu_j^A = 3, j \in \{51, 52, \ldots, 75\}$, and (iv) $\mu_j^A = 5, j \in \{76, 77, \ldots, 100\}$.

The values for $\mu_j^B, j \in \mathcal{M}$ are as follows: $\mu_j^B = \mu_j^A + 1, j \in \{1, 2, \ldots, 33\}$, (ii) $\mu_j^B = \mu_j^A + 2, j \in \{34, 35, \ldots, 67\}$, and (iii) $\mu_j^B = \mu_j^A + 1, j \in \{68, 69, \ldots, 100\}$.

1. Assume that $t_p = 25$. Note that this corresponds to the case where company X allocates teams to engines that have label "danger".
2. Solve the optimization problem given the parameter values that are provided.
3. Use clear and correct notation. Clearly define (i) the decision variables, (ii) objective function(s), (iii) constraint(s), (iv) parameters of the problem, etc.
4. In your solution you should provide: (i) a list of maintenance tasks for teams of type A and teams of type B with start-date, end-date and the machine that will be maintained; (ii) the total penalty costs; (iii) penalty costs for each machine.
5. Discuss your findings. Have you found the optimal solution?

Part C

The file `RUL_consultancy_predictions_A3.csv` contains predictions for the RUL made by a consultancy company.

1. Solve the optimization problem of the previous task (Optimization Task 1, Part B) with the predictions of the consultancy company.
2. Discuss your findings. Are the results different? If so, can you explain these differences?

Part D

1. Solve the optimization problem of the previous task (Optimization Task 1, Part B and Part C) but with a planning horizon of $t_p = 40$. Note that this corresponds to the case where company X allocates teams to engines that have label "danger" and "caution".
2. Are the engines from the previous task (Optimization Task 1, Part B and Part C) now assigned in a different way?
3. Discuss your findings. Are the results different? If so, can you explain these differences?

Optimization Task 2

In practice the working conditions of the teams also needs to be taken into account. To this end, assume that the teams have a limited number of engines that they are allowed to work on during the planning horizon. Assume that teams of type A can work on at most k_h^A engines if the planning horizon equals h. Similarly, assume that teams of type B can work on at most k_h^B engines if the planning horizon equals h. We will refer to these constraints as *max-engine constraints*.

Part A

Assume that we are currently at day $t = 1$ and that company X wants to allocate teams to engines in order to minimize penalty costs for a planning horizon of $T = t_p$. More specifically, company X wants to allocate teams to engines that have a predicted RUL that satisfies $R_j \leq T$.

1. Extend the formulation of the optimization problem of the previous task (Optimization Task 1, Part A) but this time also taking the *max-engine constraints* into account.

2. Use clear and correct notation. Clearly define (i) the decision variables, (ii) objective function(s), (iii) constraint(s), (iv) parameters of the problem, etc.

Part B

1. Assume that $k^A_{25} = 2$ and that $k^B_{25} = 2$. Solve the optimization problem of the previous task (Optimization Task 1, Part B) but this time also taking the *max-engine constraints* into account.

2. Discuss your findings. Are the results different? If so, can you explain these differences?

Part C

1. Solve the optimization problem of the previous task (Optimization Task 1, Part C) but this time also taking the *max-engine constraints* into account.

2. Discuss your findings. Are the results different? If so, can you explain these differences?

Optimization Task 3

In practice there might be additional constraints on the working conditions of the teams and the location of the engines. Assume that teams need to work close to their homes and therefore cannot travel very far to perform maintenance. In particular, we will consider the following scenario. Every engine is located in a *region* and every team can only be assigned to jobs in a particular region. We will refer to these constraints as *regional constraints*. Formally, let $L_j \in \mathcal{L}$ denote the region where engine $j \in \mathcal{M}$ is located. The indices of the regions are given by the set $\mathcal{L} = \{1, \ldots, N\}$ for some finite positive integer N. According to the *regional constraints*, a team $m \in \mathcal{G}$ can only work on an engine which is located in region $G_m \in \mathcal{L}$.

Example: There are 4 teams, 2 teams of type A and 2 teams of type B. Then $\mathcal{G}^A = \{1, 2\}$, $\mathcal{G}^B = \{3, 4\}$, $\mathcal{G} = \{1, 2, 3, 4\}$. Assume that $\mathcal{L} = \{1, 2, 3\}$. Then $L_{19} = 1$ means that engine 19 is located in region 1. Also, $G_2 = 3$ means that the second team (which has type A) can only work on engines in region 3. Also, $G_3 = 3$ means that the third team (which has type B) can only work on engines in region 3. Also, $G_4 = 2$ means that the fourth team (which has type B) can only work on engines in region 2.

Part A

Assume that we are currently at day $t = 1$ and that company X wants to allocate teams to engines in order to minimize penalty costs for a planning horizon of $T = t_p$. More specifically, company X wants to allocate teams to engines that have a predicted RUL that satisfies $R_j \leq T$.

1. Extend the formulation of the optimization problem of the previous task (Optimization Task 2, Part A) but this time also taking the *regional constraints* into account.

2. Use clear and correct notation. Clearly define (i) the decision variables, (ii) objective function(s), (iii) constraint(s), (iv) parameters of the problem, etc.

Part B

1. Assume that $\mathcal{L} = \{1, 2\}$.

   The values for $L_j, j \in \mathcal{M}$ are as follows: $L_j = 1, j \in \{1, 2, \ldots, 33\}$, and (ii) $L_j = 2, j \in \{34, 35, \ldots, 100\}$.

The values for $G_m, m \in \mathcal{G}$ are as follows: $G_m = 1, m \in \{1,3\}$, and (ii) $G_m = 2, m \in \{2,4\}$.

Solve the optimization problem of the previous task (Optimization Task 2, Part B) but this time also taking the *regional constraints* into account.

2. Discuss your findings. Are the results different? If so, can you explain these differences?

Part C

1. Solve the optimization problem of the previous task (Optimization Task 2, Part C) but this time also taking the *regional constraints* into account.
2. Discuss your findings. Are the results different? If so, can you explain these differences?

Optimization Task 4

Discuss your main findings from the previous optimization tasks. What are the managerial insights that you can provide to company X? What makes this maintenance allocation problem hard? What are easy aspects of problem?