TASK 1B

Regression (graded)

You are in the group Qwerty consisting of hongk (hongk@student.ethz.ch (mailto:///[u'hongk@student.ethz.ch'])) and kshum (kshum@student.ethz.ch (mailto:///[u'kshum@student.ethz.ch'])).

1. READ THE TASK DESCRIPTION

2. SUBMIT SOLUTIONS

3. HAND IN FINAL SOLUTION

1. TASK DESCRIPTION

This task is about **linear regression**: given an input vector **x**, your goal is to predict a value **y** as a **linear** function of a set of feature transformations, $\phi(\mathbf{x})$.

DATA DESCRIPTION

Download handout (/static/task1b_ow9d8s.zip)

In the handout for this project, you will find the the following files:

- **train.csv** - the training set
- **sample.csv** - a sample submission file in the correct format

Each line in train.csv is one data instance indexed by an Id. It consists of one float for y and 5 floats for the vector x1-x5:

```
Id,y,x1,x2,x3,x4,x5
0,-7.912551482,1.27626589,-0.854628079,1.623901111,2.14531129,2.037190468
...
```

For your convenience, we further provide a sample submission file:

```
1
2
...
```

The submission format is explained in the section below.

FEATURES DESCRIPTION

You are required to use the following features (in the following order) to make your predictions:

- Linear

$$\phi_1(\mathbf{x}) = x_1,\ \phi_2(\mathbf{x}) = x_2,\ \phi_3(\mathbf{x}) = x_3,\ \phi_4(\mathbf{x}) = x_4,\ \phi_5(\mathbf{x}) = x_5,$$

- Quadratic

$$\phi_6(\mathbf{x}) = x_1^2,\ \phi_7(\mathbf{x}) = x_2^2,\ \phi_8(\mathbf{x}) = x_3^2,\ \phi_9(\mathbf{x}) = x_4^2,\ \phi_{10}(\mathbf{x}) = x_5^2,$$

- Exponential

$$\phi_{11}(\mathbf{x}) = e^{x_1},\ \phi_{12}(\mathbf{x}) = e^{x_2},\ \phi_{13}(\mathbf{x}) = e^{x_3},\ \phi_{14}(\mathbf{x}) = e^{x_4},\ \phi_{15}(\mathbf{x}) = e^{x_5}$$

- Cosine

$$\phi_{16}(\mathbf{x}) = \cos(x_1),\ \phi_{17}(\mathbf{x}) = \cos(x_2),\ \phi_{18}(\mathbf{x}) = \cos(x_3),\ \phi_{19}(\mathbf{x}) = \cos(x_4),\ \phi_{20}(\mathbf{x}) = \cos(x_5)$$

- Constant

$$\phi_{21}(\mathbf{x}) = 1$$

where we indicate the whole input vector with **x** and we use x_i to denote its i^{th} component.

Your predictions are calculated as a linear function of the features above according to the following formula:

$$\hat{y} = w_1\phi_1(\mathbf{x}) + w_2\phi_2(\mathbf{x}) + \cdots w_{21}\phi_{21}(\mathbf{x})$$

SUBMISSION FORMAT

You are required to submit the weight of your linear predictor in a .csv file.

The file should contain 21 lines containing a float each. The i-th line indicates the i-th weight of your linear predictor:

```
1
2
...
```

The evaluation on this task depends on the accuracy of your predictor on a test data set that we do not release.

Notice that, to compute your prediction on the test data, the raw features are transformed according to the transformations introduced in the previous section. This means that the first entry of your weight vector is multiplied by $\phi_1(\mathbf{x})$, the second entry is multiplied by $\phi_2(\mathbf{x})$ and so on. As a consequence, it is important to submit the weight vector in the **correct order**.

Please keep in mind that, as a group, you have a limited number of submissions as stated on the submissions page.

EVALUATION

The evaluation metric for this task is the **Root Mean Squared Error** which is the square root of the mean/average of the square of all of the error.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

How to compute it in Matlab:

```
RMSE = sqrt(mean((y-y_pred).^2));
```

How to compute it in R:

```
RMSE <- sqrt(mean((y-y_pred)^2))
```

How to compute it in Python:

```
from sklearn.metrics import mean_squared_error
RMSE = mean_squared_error(y, y_pred)**0.5
```

GRADING

We have partitioned the test set into two parts and use it to compute a *public* and a *private* score for each submission. You only receive feedback about your performance on the public part in the form of the public score, while the private leaderboard remains secret. The purpose of this division is to prevent overfitting to the public score. You are motivated to make sure your models will generalize well even to the private part of the test set.

When handing in the task, you need to select which of your submissions will get graded and provide a short description of your approach. We will then compare your selected submission to three baselines (easy, medium and hard). Your final grade depends on both the public score and the private score (weighted equally) and on a properly-written description of your approach. The following **non-binding** guidance provides you with an idea on what is expected to obtain a certain grade: If you hand in a properly-written description and your handed-in submission performs better than the easy baseline, you may expect a grade exceeding a 4. If it further beats the medium baseline, you may expect that the grade will exceed a 5. If in addition your submission performs equal to or better than the hard baseline, you may expect a 6. If you do not hand in a properly-written description of your approach, you may obtain zero points regardless of how well your submission performs.

⚠ Make sure that you properly hand in the task, otherwise you may obtain zero points for this task.

FREQUENTLY ASKED QUESTIONS

WHICH PROGRAMMING LANGUAGE AM I SUPPOSED TO USE? WHAT TOOLS AM I ALLOWED TO USE?

You are free to choose any programming language and use any software library.

CAN YOU HELP ME SOLVE THE TASK? CAN YOU GIVE ME A HINT?

As the tasks are a graded part of the class, **we cannot help you solve them**. However, feel free to ask general questions about the course material during or after the exercise sessions.

CAN YOU GIVE ME A DEADLINE EXTENSION?

> ⚠ We do not grant any deadline extensions!

CAN I POST ON PIAZZA AS SOON AS HAVE A QUESTION?

This is highly discouraged. Instead,

- Read the details of the task thoroughly.
- Review the frequently asked questions.
- If there is another team that solved the task, spend more time thinking.
- Discuss it with your team-mates.

If you still consider that you should contact the TAs, you can post a **private** question on Piazza. Remember that collaboration with other teams is prohibited.

WHEN WILL I RECEIVE THE PRIVATE SCORES? AND THE PROJECT GRADES?

Before the exam, you will obtain an email with all your private scores. We do not release scores before that. The same email will also contain your project grade.