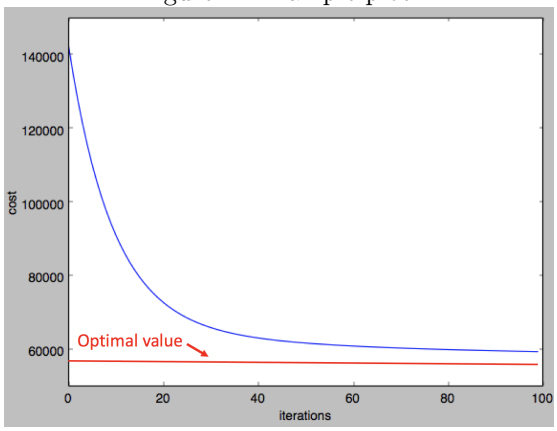
Homework 1: Linear Regression and Logistic Regression

Dataset. In this programming homework, we will use two LIBSVM datasets which are pre-processed data originally from UCI data repository.

- Linear regression - Housing dataset (We will use housing_scale dataset). Predict housing values in suburbs of Boston. `https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/regression/housing_scale`.
- Logistic regression - Adult dataset (We will only use a3a training dataset). Predict whether income exceeds \$50K/yr based on census data. `https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/binary/a3a`.

1. *Linear regression.* Randomly split the dataset into two groups: training (around 80%) and testing (around 20%). Learn the linear regression model on the training data, using the analytic solution. Compute the prediction error on the test data: $\frac{1}{\#\text{of test points}} \sum_{i \in \text{testset}} |\hat{y}_i - y_i^*|$ where $\hat{y}_i$ and y_i^* are the prediction and the true value respectively for data point i. Repeat this process 10 times and report all individual prediction errors of 10 trials and the average of them.
2. *Linear regression.* Do the same work as in the problem #1 but now using a gradient descent. (10 randomly generated datasets in #1 should be maintained; we will use the datasets generated in #1.) Here we are not using (exact or backtracking) line searches. You need to try several selections for the *fixed* step size.
   (a) Compare prediction errors with those from #1.
   (b) Additionally draw plots showing objective function values vs. iterations of gradient descent. In all plots, optimal objective function value by analytic solution should be presented (it would be a horizontal line, as shown in Figure 1). Report cases for different step sizes (for too large, proper and too small step sizes).
3. *Logistic regression.* As in the problem #1, randomly split the adult dataset into two groups (80% for training and 20% testing). Learn logistic regression on the training data. Here we compare the performances of gradient descent methods i) with fixed-sized step sizes and ii) with the backtracking line search. Try to find the best step size for i) and the best hyperparameters α and β for ii) (in terms of the final objective function values).

Figure 1: Example plot.



(a) Compare objective function values vs. iterations of gradient descent by approaches i) and ii) with their best choices.

(b) Report (the individual and the average over 10 trials) prediction errors on the test data: $\frac{1}{\#\text{of test points}} \sum_{i \in \text{testset}} \mathcal{I}(\hat{y}_i \neq y_i^*)$ where $\mathcal{I}(a)$ is the indicator function (1 if a is true and 0 otherwise) by approaches i) and ii) with the best choices.

4. Provide the report with the details. (in Korean or in English; either is fine.) Try to include discussions and your impressions on your work.