Introduction to Deep Learning for Speech and Text Processing
Exercise Sheet 4: Machine Learning - Basics

Thang Vu

7th November 2025

Clustering

Exercise 1.

In speaker recognition or speech synthesis the information from speakers is often represented in high dimensional *speaker embedding vectors* $\in \mathbb{R}^n$. The *speaker embedding vectors* are computed on utterance level, i.e. we compute a single vector for each utterance. Now, given a set of utterances from two different speakers, we want to evaluate how well our embedding function performs.

The reduced speaker embedding vectors on utterance level are as follows:

$$v(0) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, v(1) = \begin{bmatrix} 1 \\ 1.5 \end{bmatrix}, v(2) = \begin{bmatrix} 1.5 \\ 1.5 \end{bmatrix}, v(3) = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, v(4) = \begin{bmatrix} 2.5 \\ 3 \end{bmatrix}, v(5) = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, v(6) = \begin{bmatrix} 2.5 \\ 1 \end{bmatrix}, v(7) = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, v(8) = \begin{bmatrix} 2.5 \\ 1.5 \end{bmatrix}$$

Perform two update steps using the k-means algorithm with $k = 2$ clusters. The initial means are $v(1)$ and $v(7)$. Use euclidean distance for this task and assign every data point to its closest centroid.

Features

Exercise 2.

The primary goal of sentiment analysis is to understand and categorize the opinions expressed in texts. It can be considered as a classification task in machine learning. In a traditional machine learning pipeline, the following features are reasonable:

(1) Word features: the average word embedding of all words of a sentence

(2) Aggregated sentiment scores: the sum of all sentiment scores of each word

(3) Last word sentiment: the sentiment score of the last word

Given the following word embeddings:

$$v('it') = \begin{bmatrix} 0.5 \\ 0.5 \\ 0 \end{bmatrix}, v('this') = \begin{bmatrix} 0.4 \\ 0.6 \\ -0.7 \end{bmatrix}, v('is') = \begin{bmatrix} 0 \\ 0 \\ 0.8 \end{bmatrix}, v('movie') = \begin{bmatrix} 0.2 \\ 0.4 \\ -0.8 \end{bmatrix}, v('a') = \begin{bmatrix} 0.2 \\ -0,2 \\ 0 \end{bmatrix}$$

$$v('good') = \begin{bmatrix} 0.7 \\ 0 \\ 0 \end{bmatrix}, v('bad') = \begin{bmatrix} -0.8 \\ 0 \\ 0.1 \end{bmatrix}, v('great') = \begin{bmatrix} 0.9 \\ 0.1 \\ 0 \end{bmatrix}, s('not') = \begin{bmatrix} -0.2 \\ -0.2 \\ 0.1 \end{bmatrix}$$

and the following sentiment dictionary:

$$s('it') = 0, s('this') = 0, s('is') = 0, s('movie') = 0.2, s('a') = 0$$

$$s('good') = 0.8, s('bad') = -0.6, v('great') = 0.7, s('not') = -0.2$$

- extract the aforementioned features and construct the feature vector of the following inputs:
  (1) it is a great movie
  (2) this movie is not a good movie
  (3) it is a bad movie
- which features have low variance and might not be useful based on the above examples?

Logistic Regression

Exercise 3.
Compute the predictions for the following logistic regression model $p(y = 1|x)$ and the following test dataset. The model is trained to binary-classify sentences: a label of 1 refers to a positive sentiment, a label of 0 to a negative sentiment.

Logistic Regression Model:

$$w = \begin{bmatrix} -2 \\ 4 \\ -1 \\ 0.5 \end{bmatrix}, \quad b = -1$$

Dataset:

$$(x_1 = \begin{bmatrix} -1 \\ 2 \\ 0 \\ -2 \end{bmatrix}, \quad y_1 = 1), \quad (x_2 = \begin{bmatrix} 2 \\ 0 \\ 2.5 \\ 2 \end{bmatrix}, \quad y_2 = 0), \quad (x_3 = \begin{bmatrix} 1 \\ 1 \\ 2 \\ 0 \end{bmatrix}, \quad y_3 = 1) \quad (x_4 = \begin{bmatrix} 1 \\ 0.5 \\ -1 \\ 0 \end{bmatrix}, \quad y_4 = 1)$$

How good is the model in your opinion? Explain your intuition.

Learning

Optimization

Exercise 4.
Given the following linear regression model and the following training dataset.

Model:

$$w = \begin{bmatrix} 3 \\ -1.5 \\ 2 \\ 0.5 \end{bmatrix}, \quad b = -1$$

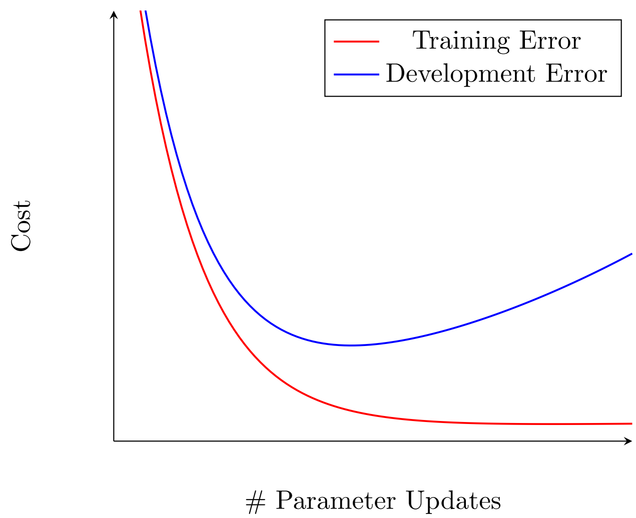
Dataset:

$$(x_1 = \begin{bmatrix} 0 \\ -1 \\ 1 \\ 4 \end{bmatrix}, \quad y_1 = 3), \quad (x_2 = \begin{bmatrix} 2 \\ -1 \\ -2 \\ -1 \end{bmatrix}, \quad y_2 = -1)$$

(1) Calculate the mean squared error of the dataset.
(2) Compute the derivative of the error with respect to its weights w and b.
(3) Perform one step of gradient descent with a learning rate of $\eta = 0.1$.

Training curves

Exercise 5.
When training a ML model and monitoring its loss on the training and development data, the following curves are observed:



Do the curves behave as expected? Explain your answer.

General questions

Exercise 6.
Mark whether the following statements are true or false and explain in short your solutions.

Statement	**True**	**False**
Hyperparameters are tuned to optimize the results on the training set.		
An overfitted model perfectly matches the development data.		
Support vector machines can also handle multiclass classification tasks.		
Decision tree classifiers are easy to interpret and visualize		
In order to train LDA, we need labelled data.		