

機械学習基礎 2B 課題

締切 : 8 月 12 日 (火) 17:00

回帰に関する以下の問いに答えよ.

1. 線形の次元削減の概略および主成分分析について説明せよ.
2. データファイル (iris.txt) をダウンロードして, 主成分分析によって 2 次元平面に写像するプログラムを作成せよ.
データファイルは, アヤメのデータと呼ばれるもので, 3 種類のアヤメが各種 50 サンプルずつ (計 150 サンプル) 含まれ, 各サンプルは, 「がく片の長さ」, 「がく片の幅」, 「花びらの長さ」, 「花びらの幅」の 4 つの特徴量で表現されている.
データは, 上から “setosa” 種 50 サンプル, “versicolor” 種 50 サンプル, “virginica” 種 50 サンプルが含まれている.
3. 写像した 2 次元平面の各軸がどのような特徴量で構成されているかについて述べよ. また, 各軸の寄与度について説明せよ.
4. 上記の結果を踏まえて, 次元削減の有用性について議論せよ.

Assignment for Fundamental Machine Learning 2B

Deadline: 5pm, 12 August

Answer the following questions about regression.

1. Explain the outline of linear dimensionality reduction and principal component analysis.
2. Download the data file (iris.txt) and create a program to map the data onto a two-dimensional plane by principal component analysis.
The data file is called iris data and contains 50 samples of each of the three types of iris (150 samples in total), and each sample is represented by four features: "sepal length," "sepal width," "petal length," and "petal width."
The data includes 50 samples of "setosa" species, 50 samples of "versicolor" species, and 50 samples of "virginica" species from the top.
3. Describe what kind of features each axis of the mapped two-dimensional plane consists of. Explain the contribution of each axis.
4. Based on the above results, discuss the usefulness of dimensionality reduction.