

Варианты лабораторной работы №3 по курсу «Машинное обучение. Часть 1»

Файл с данными

data_vN-NN.csv, N-NN – номер варианта

Шаблон отчета

https://docs.google.com/presentation/d/1SaG1JYkH_X5eFrq6VleHDd757jmKna8wxb3YyGyK7h0/edit?usp=sharing

Отчет

Файл otchet_vN-NN_GroupFIO.pdf и **исходные коды**, написанные для выполнения заданий, высылать на mlearninglabs@gmail.com

Указания по оформлению графиков

<http://datalearning.ru/study/Courses/methodic/lections/graphs.pdf>

К каждому графику в отчете должны быть даны все необходимые пояснения для его понимания (название, подписи осей, легенда, условия эксперимента, параметры обработки и пр.).

Рекомендуемые средства программной реализации

Python, MATLAB

Литература

1. Alpaydin, E. (2014). *Introduction to machine learning*. MIT press.
2. Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms*. Cambridge university press.
3. Raschka, S. (2015). *Python machine learning*. Packt Publishing Ltd.

Вариант 3-06

Kernel regression

Задания

1. Построить регрессограммы для различного числа интервалов группировки: а) 3; б) 10; в) 50. Построить график зависимости MSE непараметрической регрессии на обучающей, валидационной и тестовой выборках от числа интервалов группировки.
2. Построить графики зависимости MSE ядерной регрессии на обучающей, валидационной и тестовой выборках от ширины окна для: а) прямоугольного окна; б) гауссова окна; в) окна Епанечникова; г) треугольного окна. Определить оптимальную ширину каждого окна.
3. Визуализировать графики ядерной функции регрессии, полученной с использованием: а) прямоугольного окна; б) гауссова окна; в) окна Епанечникова; г) треугольного окна. Ширину окна задать: i) равной соответствующему оптимальному значению, полученному в п.2; ii) значению, меньшему оптимального; iii) значению, большему оптимального.
4. Построить графики зависимости дисперсии и смещения регрессионной модели для нескольких фиксированных значений x от ширины окна для: а) прямоугольного окна; б) гауссова окна; в) окна Епанечникова; г) треугольного окна.
5. Построить график зависимости MSE KNN-регрессии от числа ближайших соседей K на обучающей и тестовой выборках. Определить оптимальное число ближайших соседей и визуализировать графики KNN-регрессии для различных значений K : а) равному оптимальному; б) меньшего оптимального; в) большего оптимального.
6. Сделать выводы о влиянии вида и ширины окна на качество ядерной регрессионной модели.

Указания

Разбиение на обучающую, валидационную и тестовую выборки – holdout в отношении 60/30/10.

В пп.2, 5 оптимальные значения гиперпараметров обучения определять по значению ошибки на валидационной выборке.

Литература

1. Hansen, B. E. (2009). Lecture notes on nonparametrics. *Lecture notes*.
2. Fan, J., and I. Gijbels. (1996). *Local Polynomial Modelling and Its Applications*. Vol. 66. Monographs on Statistics and Applied Probability. London: Chapman & Hall.
3. Chu, C. K., & Marron, J. S. (1991). Choosing a kernel regression estimator. *Statistical science*, 404-419.
4. Wand, M. P., and M. C. Jones. (1995). *Kernel Smoothing*. Vol. 60. Monographs on Statistics and Applied Probability. London: Chapman & Hall.