Лабораторная работа №2.

1. Загрузить данные. В работе используются данные Walmart.csv.

Это данные о продажах с 05.02.2010 по 01.11.2012 в сети магазинов Walmart. Состав данных:

номер магазина

неделя продаж

продажи для данного магазина

является ли неделя особым праздником неделя 1 – выходная неделя 0 – непраздничная неделя

Температура в день продажи

Стоимость топлива в регионе

Преобладающий индекс потребительских цен

Преобладающий уровень безработицы

Праздничные мероприятия в США:

Суперкубок: 12 февраля 2010 г., 11 февраля 2011 г., 10 февраля 2012 г., 8 февраля 2013 г.

День труда: 10-сен-10, 9-сен-11, 7-сен-12, 6-сен-13.

День Благодарения: 26 ноября 10, 25 ноября 11, 23 ноября 12, 29 ноября 13.

Рождество: 31 декабря 2010 г., 30 декабря 2011 г., 28 декабря 2012 г., 27 декабря 2013 г.

Целевым признаком является Weekly\_Sales - продажи для данного магазина

2. Импортировать из библиотеки sklearn все необходимые библиотеки машинного обучения и анализа данных.

3. Обработать данные (если это необходимо):

привести все данные к числовой форме;

отбросить ненужные данные;

проверить, есть ли пропуски;

разделить данные на обучающую и тестовую выборки.

составить матрицу корреляции и удалить коррелирующие признаки, если они есть.

4. Трактуя задачу о как задачу регрессии, построить на очищенных данных многослойный персептрон, предсказывающий продажи для магазинов.

5. Рассчитаете следующие метрики, чтобы проверить качество сети:

- коэффициент детерминации (или коэффициент R2):

- RMSE (root mean squared error);

- среднее абсолютное отклонение (mean absolute error, MAE);

- среднеквадратичная логарифмическая ошибка (mean squared logarithmic error, MSLE).

6. Построить графики потерь на этапах обучения и проверки