Задание 2.

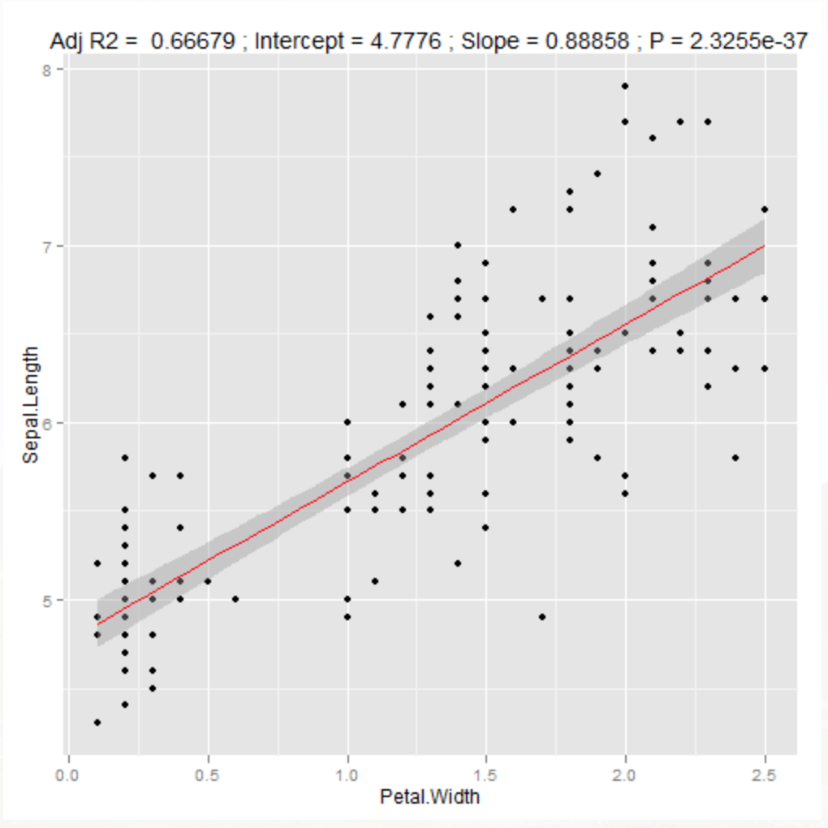
Введение в машинное обучение на языке R

В этом задании Вам необходимо создать R-скрипт, в котором реализовать модель линейной регрессии. Для установки R необходимо перейти на узел:

<https://cran.r-project.org/bin/windows/base/>

Скрипт должен содержать:

1. Чтение данных формате csv. (Необходимо найти файл с данными, в котором реализовать их перечисление через знак запятой).
2. Обучение модели и нахождение коэффициентов.
3. Вывод результатов на экран.
4. Графическое изображение исходных данных и найденной прямой линии.

Приблизительный вид результатов приведен на рисунке слева. После отладки скрипта необходимо написать отчет, в котором ответить на следующие вопросы:

1. Описать модель линейной регрессии и вывести соответствующие математические формулы. Приведите пример задачи, которую можно изучать с помощью предлагаемых методов.
2. Откуда были взяты Ваши данные и что они означают? Здесь описать, для чего используется модель линейной регрессии.
3. Какие трудности возникли у Вас при написании скрипта. Здесь указать как Вы загружали данные (функция read.csv), как оценивалась модель (функция lm), как был организован вывод результатов (функция summary), изображение данных (plot) и прямой линии (а какая функция рисует линию?).
4. Дать возможные рекомендации для усовершенствования и усложнения изучаемой модели (рассмотреть, скажем, модели множественной регрессии, логической регрессии и пр.).
5. Рассмотреть в первом приближении более сложные алгоритмы машинного обучения (скажем методы наивного байесовского классификатора, опорных векторов, кластеризации и пр.). Здесь достаточно минимального ответа, особенно в случае, если Вы плохо поняли эту теорию.