CS342 Machine Learning: Worksheet #2

Seminar in Week 2 of Term 2

Week 16

Office Hours:
CS 3.07, Mondays & Fridays 16:00-17:00

Instructor:
Dr Theo Damoulas (T.Damoulas@warwick.ac.uk)

Tutors:
Helen McKay (H.McKay@warwick.ac.uk),
Joe Meagher (J.Meagher@warwick.ac.uk)
Karla Monterrubio-Gomez (K.Monterrubio-Gomez@warwick.ac.uk),
Jevgenij Gamper (J.Gamper@warwick.ac.uk)

The Worksheet builds on Chapter 1 of the recommended textbook *A First Course in Machine Learning* (2nd Edition, CRC Press) and the first 3 lectures of CS342 on linear regression and the Ordinary Least Squares (OLS) solution.

Problem 1

[Q1.10 from R&G book]:

Derive the optimal least squares parameter value, $\hat{\mathbf{w}}$, for the total training loss

$$\mathcal{L} = \sum_{n=1}^{N}(t_n - \mathbf{w}^{\mathrm{T}}\mathbf{x}_n)^2$$

How does the expression compare with that derived from the average loss?

Problem 2

[Q1.11 from R&G book]:

The following expression is known as the *weighted* average loss:

$$\mathcal{L} = \frac{1}{N}\sum_{n=1}^{N}\alpha_n(t_n - \mathbf{w}^{\mathrm{T}}\mathbf{x}_n)^2$$

where the influence of each data point is controlled by its associated α parameter. Assuing that each α_n is fixed, derive the optimal least squares paramater value $\hat{\mathbf{w}}$

Problem 3

In linear regression, if instead of the squared error $(t_n - \mathbf{w}^{\mathrm{T}}\mathbf{x}_n)^2$ we use the absolute error $|t_n - \mathbf{w}^{\mathrm{T}}\mathbf{x}_n|$ as a loss function we have a model that is called *Least Absolute Deviations* and is more robust to outliers then Ordinary Least Squares (OLS).

Why would this loss function lead to a model that is more robust to outliers? Attempt to derive the optimal parameter values for this loss function. What is the problem and the drawback in comparison to OLS?