CS342 Machine Learning: Worksheet #1 Solution

Seminar in Week 1 of Term 2

Week 15

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 First Worksheet reviews introductory material in linear algebra, random variables, and probability theory that will be needed for the module and the following worksheets.

The questions listed point to key concepts that you can discuss and review with your tutors.

A recommended reference for the material covered in this worksheet is Deep Learning. The material in Question 1 is covered in the chapter on Linear Algebra, while Question 2 is covered in the chapter on Probability and Information Theory. For the most part, these solutions will refer to this text, with the relevant section referenced in blue. The solution will also be presented in blue.

Question 1

Some basic linear algebra:

- What is a scalar $x \in \Re$, a vector $\mathbf{x} \in \Re^D$, a matrix $\mathbf{X} \in \Re^{N \times N}$ etc. *Note the bold and capital fonts used to denote the multiple dimensions.* (2.1)
- What is the transpose of a vector $\mathbf{x}^{\mathrm{T}}$, the trace of a matrix trace($\mathbf{X}$), the inverse of a matrix $\mathbf{X}^{-1}$ the determinant of a matrix $\det(\mathbf{X}) = |\mathbf{X}|$. Assume all can be defined etc. (2.1), (2.3), (2.10), (2.11)
- Show that

$$\mathbf{w}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{w} = w_1^2 \left(\sum_{n=1}^N x_{n1}^2\right) + 2w_1w_2 \left(\sum_{n=1}^N x_{n1}x_{n2}\right) + w_2^2 \left(\sum_{n=1}^N x_{n2}^2\right)$$

  where $\mathbf{w} \in \Re^D$ is a column vector, $\mathbf{X} \in \Re^{N \times D}$ is a matrix and $D = 2$.

  Let $\mathbf{w}^{\mathrm{T}} = [w_0 \;\; w_1]$ and $\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ \vdots & \vdots \\ x_{N1} & x_{N2} \end{bmatrix}$

  Then $\mathbf{X}^{\mathrm{T}}\mathbf{X}$, then $\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{w}$, finally $\mathbf{w}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{w}$ gives the solution.

- Show that: a) the dot product between two vectors is *commutative*, b) matrix multiplication is *associative*, c) matrix multiplication is *distributive*, d) matrix multiplication is **not** *commutative.* (2.2)

  Try with examples to convince yourself.

- What is: a) the inverse of the identity matrix $\mathbf{I}$, b) the inverse of $\phi\mathbf{I}$ where $\phi \in \Re$, c) the relationship between the inner (dot) product and the angle of two vectors $\mathbf{x}, \mathbf{z} \in \Re^D$. Equation (2.34)

Question 2

Some basic probability theory

- What is the difference of discrete and continuous random variables (RVs)? Give some examples. (3.2)
- When does an RV have a probability mass function (pmf) and when a probability density function (pdf)? (3.3)
- if X is independent of Y, what does $P(X|Y)$ equal to?

  If $X \perp Y), then P(X|Y) = P(X)$

- if X and Y are now conditionally independent given Z, what does $P(X,Y|Z)$ equal to? (3.7)
- Give the sum and product rule (or chain rule) of probability (3.4), (3.6)
- Use the chain rule to get Bayes rule.

  Recall the Bayes rule can be expressed in terms of the random variables X and Y as

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$$

Question 2 continued on next page. . .

- Define the expectation of a discrete and of a continuous random variable. What is the expected value of rolling a fair dice? (3.8)

- Define the variance of a random variable and the covariance of two random variables. (3.8)

- Show that

  $\mathrm{Var}(a + bX) = b^2\mathrm{Var}(X)$ and that if X and Y are independent $\mathrm{Var}(X + Y) = \mathrm{Var}(Y) + \mathrm{Var}(X)$ (3.9)

  Try $Var(a{+}bX) = E\left[\left((a + bX) - E\left[(a + bX)\right]\right)^2\right]$ and $Var(X{+}Y) = E\left[\left((X + Y) - E\left[(X + Y)\right]\right)^2\right]$

- Give the probability density function of the univariate Gaussian (a.k.a Normal) distribution

  For $X \sim \mathcal{N}(\mu, \sigma)$ the pdf of X is

$$p(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- What is the expected value of X if $X \sim \mathcal{N}(0, 1)$?

  $E[X] = 0$

If you have any trouble with the above set of questions, I *strongly* advice you to quickly revise some basic linear algebra and probability theory and reach out to your tutors. Some material is linked in the module website and you can also ask your Tutors for more suggestions and material.