mathematical methods - week 15

(Non)linear dimensionality reduction

Georgia Tech PHYS-6124

Homework HW #15 due Monday, December 2, 2019

== show all your work for maximum credit,
== put labels, title, legends on any graphs
== acknowledge study group member, if collective effort
== if you are LaTeXing, here is the source code

Exercise 15.1 *Unbiased sample variance* 5 points

Bonus points
Exercise 15.2 *Standard error of the mean* 5 points
Exercise 15.3 *Bayesian statistics*, by Sara A. Solla 10 points

Total of 10 points = 100 % score. Extra points accumulate, can help you still if you had missed a few problems.

edited November 24, 2019

Week 15 syllabus Monday, November 25, 2019

Linear and nonlinear dimensionality reduction:
applications to neural data
Lecturer: Sara A. Solla

Mon Neural recordings; Principal Components Analysis (PCA); Singular Value Decomposition (SVD); ISOMAP nonlinear dimensionality reduction; Multidimensional scaling

- Sara's lecture notes.
- Predrag's summary of key concepts for a physicist: ChaosBook Sect. 17.1.3 Moments, cumulants.

15.1 Optional reading: Bayesian statistics

Sara A. Solla

Natural sciences aim at abstracting general principles from the observation of natural phenomena. Such observations are always affected by instrumental restrictions and limited measurement time. The available information is thus imperfect and to some extent unreliable; scientists in general and physicists in particular thus have to face the task of extracting valid inferences from noisy and incomplete data. Bayesian probability theory provides a systematic framework for quantitative reasoning in the face of such uncertainty.

In this lecture (not given in the Fall 2019 course) we will focus on the problem of inferring a probabilistic relationship between a dependent and an independent variable. We will review the concepts of joint and conditional probability distributions, and justify the commonly adopted Gaussian assumption on the basis of maximal entropy arguments. We will state Bayes' theorem and discuss its application to the problem of integrating prior knowledge about the variables of interest with the information provided by the data in order to optimally update our knowledge about these variables. We will introduce and discuss Maximum Likelihood (ML) and Maximum A Posteriori (MAP) for optimal inference. These methods provide a solution to the problem of specifying optimal values for the parameters in a model for the relationship between independent and dependent variables. We will discuss the general formulation of this framework, and demonstrate that it validates the method of minimizing the sum-of-squared-errors in the case of Gaussian distributions.

- A quick but superficial read: Matthew R. Francis, *So what's all the fuss about Bayesian statistics?*
- Reading: Lyons [1], *Bayes and Frequentism: a particle physicist's perspective* (click here)

References

[1] L. Lyons, "Bayes and Frequentism: a particle physicist's perspective", Contemporary Physics **54**, 1–16 (2013).

Exercises

15.1. **Unbiased sample variance.** Empirical estimates of the mean $\hat{\mu}$ and the variance $\hat{\sigma}^2$ are said to be "unbiased" if their expectations equal the exact values,

$$\mathbb{E}[\hat{\mu}] = \mu\,, \qquad \mathbb{E}[\hat{\sigma}^2] = \sigma^2\,. \tag{15.1}$$

(a) Verify that the empirical mean

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} a_i \tag{15.2}$$

is unbiased.

(b) Show that the naive empirical estimate for the *sample variance*

$$\bar{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}(a_i - \hat{\mu})^2 = \frac{1}{N}\sum_{i=1}^{N} a_i^2 - \frac{1}{N^2}\left(\sum_{i=1}^{N} a_i\right)^2$$

is biased. Hint: note that in evaluating $\mathbb{E}[\cdots]$ you have to separate out the diagonal terms in

$$\left(\sum_{i=1}^{N} a_i\right)^2 = \sum_{i=1}^{N} a_i^2 + \sum_{i\neq j}^{N} a_i a_j\,. \tag{15.3}$$

(c) Show that the empirical estimate of form

$$\hat{\sigma}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(a_i - \hat{\mu})^2\,, \tag{15.4}$$

is unbiased.

(d) Is this empirical sample variance unbiased for any finite sample size, or is it unbiased only in the $n \to \infty$ limit?

Sara A. Solla

15.2. **Standard error of the mean.**

Now, estimate the empirical mean (15.2) of observable a by $j = 1, 2, \cdots, N$ attempts to estimate the mean $\hat{\mu}_j$, each based on M data samples

$$\hat{\mu}_j = \frac{1}{M}\sum_{i=1}^{M} a_i\,. \tag{15.5}$$

Every attempt yields a different sample mean.

(a) Argue that $\hat{\mu}_j$ itself is an idd random variable, with unbiased expectation $\mathbb{E}[\hat{\mu}] = \mu$.

(b) What is its variance

$$\mathrm{Var}[\hat{\mu}] = \mathbb{E}[(\hat{\mu} - \mu)^2] = \mathbb{E}[\hat{\mu}^2] - \mu^2$$

as a function of variance expectation (15.1) and N, the number of $\hat{\mu}_j$ estimates? Hint; one way to do this is to repeat the calculations of exercise 15.1, this time for $\hat{\mu}_j$ rather than a_i.

(c) The quantity $\sqrt{\mathrm{Var}[\hat{\mu}]} = \sigma/\sqrt{N}$ is called the *standard error of the mean* (SEM); it tells us that the accuracy of the determination of the mean μ. How does SEM decrease as the N, the number of estimate attempts, increases?

Sara A. Solla

15.3. **Bayes.** *Bayesian statistics*.