Homework 5

Linear regression

In the page 47 of the PowerPoint, the posterior is a normal pdf

1. To show the posterior is a normal pdf, write a computer program that plot the posterior $\dfrac{p(w|\alpha)\prod_{n=1}^{N}p(y_n|w,\beta)}{\int\; p(w|\alpha)\prod_{n=1}^{N}p(y_n|w,\beta)dw}$. All the terms are defined in the PowerPoint. You need to find data by yourself.

2. Find analytically $E_{p(w|D,\alpha^{(t)},\,\beta^{(t)})}(\ln(p(w,\,D|\,\alpha,\,\beta)))$. You must give all the details.

Approximation

Let us consider a beta pdf

1. Build the posterior. You must consider the beta pdf where the parameters are the mean and the variance (see Wikipedia). The pdf of the mean and the variance are

   $p(\mu|\sigma) \propto \sigma^{1/2}\exp(-\sigma_0\sigma(\mu-\mu_0)^2/2)$

   $p(\sigma) \propto \sigma_0^{\alpha-1}\exp(-b_0\sigma)$

2. Propose an approximation of the posterior based on variational factorization.
3. By using the data you generate yourself study the effects of the hyperparameters on both the posterior and its approximation.

Feature selection

Revisit the LEC algorithm by introducing the feature selection presented in the lecture. The algorithm should be explained clearly. Recall that it is based on the paper S. Boutemedjet, N. Bouguila, D. Ziou: A Hybrid Feature Extraction Selection Approach for High-Dimensional Non-Gaussian Data Clustering. IEEE TPAMI, 31(8): 1429-1443, 2009.

Principal component analysis

Use an existing computer program of the principal component analysis to compare the following configurations:

1. Covariance matrix is $E((x-m)(x-m)^t)$, where $E()$ is the expectation, x is a vector.
2. Non-centered covariance matrix $E(x\,x^{\wedge}t)$

You need to generate the data or to find them on the Web.

Random process

1. Let us consider that students arriving to SIAT can be modeled by using Poisson process with a constant frequency A. Find the probability that the number of students arriving during the time interval [t_{i-1}, t_i] is equal to 10.
2. Let consider the discrete time random process $Y_t = \mu t + \varepsilon_t$, where t=1,2... Find the covariance and the autocorrelation.

Deliverable: the answers and the source codes you write by yourself.

Deadline: June 3rd, 2019