Indian Institute of Space Science and Technology

Department of Avionics

Course Name: Pattern Recognition and Machine Learning (PRML)

PRML Programming Assignment 3

---

1. This assignment is due on: **14 /11/2020**
2. This assignment should be submitted either as an **ipython** notebook or a **Latex** compiled document with **codes, results and conclusion**.
3. Use python 3 for implementing.
4. Use **Numpy** (for array processing and linear algebra operations), **Matplotlib** (for Plotting)
5. For reading the data, **pandas** or **python3 inbuilt** file operations can be used.
6. Data files are provided with the link, click on the link text to download the data

---

Question 1: (Parametric Density estimation: MLE)

1. Generate a data from a normal distribution of 1000 points, with mean 50 and standard deviation 5.
2. Now, pretend that you do not know the underlying distribution of the above data; Plot a histogram for the 1000 data points, with no. of bins equal to 10.
3. Assuming a Gaussian normal distribution, estimate the parameters of the distribution using MLE and plot the expected distribution over the histogram plot.

Question 2: (Parametric Density estimation: MAP)

Consider Bayesian estimation of the mean of a one-dimensional Gaussian. Suppose you are given the prior for the mean as $p(\mu) \sim \mathcal{N}(\mu_0, \sigma_0)$.

D (Data) = {-0.014, 0.48, 0.32, 1.4, 3.1, 0.11, 0.14, 2.2, -0.46, -0.49}

(a) Write a program that plots the density $p(x|D)$ given μ_0, σ_0, σ and training set $D = \{x_1, x_2, \dots, \dots, x_n\}$.

(b) Estimate μ for the given data; assume $\mu_0 = -1$ and plot your estimated densities $p(x|D)$ for each of the following values of the dogmatism, $\frac{\sigma^2}{\sigma_0^2}$: 0.1, 1.0, 10, 100.

Question 3: (Kernel density estimation)

I. Consider a one dimensional two class classification problem, where we have collected the following data for each class: D1 = {-3,-2,8,3,6,7} and D2 = {-5,-2,3,4,8}. Suppose we decided to use Parzen windows with window width h = 2 and ϕ (x) defined as:

$$\phi(u) = \begin{cases} 4e^{-4u}; \forall\, u > 0 \\ 0; \;\; else \end{cases}$$

(a) Classify sample x = 4 using the ML classifier

(b) Plot the density estimate $p_\phi\ (x)$ for $D_1\ and\ D_2$

II.

1. Generate samples from 2 different Gaussian distributions, with means 20 and 40 respectively, & standard deviations of 5. With 300 samples from distribution the first distribution and 700 samples from the second. (Finally stack the data points together; this are the samples that we use)
2. For the above generated samples plot the histogram, no. of bins = 50
3. Perform the kernel density estimation with Gaussian kernel of bandwidth 2. (For this you may use "scikit learn's kernelDensity " : `sklearn.neighbors.KernelDensity`)
4. Plot the estimated density. Also estimate and plot the estimated density for different bandwidths. (e.g. for 1, 5 and 10)
5. Report your conclusions.