

فرمت گزارش:

گزارش باید به زبان فارسی و در قالب فایل PDF باشد. در گزارش، تحلیل و نتیجه‌گیری خود را در رابطه با هر تمرین به طور مختصر، در حد یک پاراگراف، بیان نمایید.

فایل گزارش خود را به شکل Report2_StdNum.pdf نامگذاری نمایید (مانند Report2_8931064.pdf)

فرمت کدها:

برای هر تمرین باید فایل کد جداگانه‌ای در محیط MATLAB ، R یا Python تهیه شود. هر فایل کد خود را به شکل CE2_ProblemNum_StdNum نامگذاری کنید.

نحوه تحویل:

فایل‌های کد و گزارش خود را که طبق فرمت‌های فوق تهیه شده‌اند، در قالب یک فایل فشرده در سایت درس بارگذاری نمایید. فایل فشرده را به شکل CE2_StdNum نامگذاری نمایید.

مهلت ارسال تمرین ساعت 23:55 دقیقه‌ی روز شنبه ۲ دی می‌باشد.

ضمناً به ازای هر روز تاخیر در ارسال تمرین، 10 درصد از نمره‌ی آن کم می‌شود.

هر گونه سوال در مورد تمرین را میتوانید از طریق ایمیل به آدرس z_naraghi@aut.ac.ir ارسال نمایید.

1. The Kullback Leibler divergence is not symmetric, so the solutions of $\operatorname{argmin}_q D_{KL}(p||q)$ and $\operatorname{argmin}_q D_{KL}(q||p)$ are different. Consider p as a simple Gaussian mixture model: sum of $\text{Normal}(3,0.5)$ and $\text{Normal}(7,0.5)$. Find the $q1^* = \operatorname{argmin}_q D_{KL}(p||q)$ and $q2^* = \operatorname{argmin}_q D_{KL}(q||p)$ among three pdfs: $q \sim \text{Normal}(3,0.5)$, $q \sim \text{Normal}(5,1.5)$ and $q \sim \text{Normal}(7,0.5)$. Plot $q1^*$ and $q2^*$ on p distribution. In image generating field, consider p as a model of image samples and $q1^*$ and $q2^*$ as the learned model. What's the difference between images which are generated from $q1^*$ and $q2^*$?
2. Generate 100 observations from a $\text{Normal}(0,1)$ distribution. Compute a 95 percent confidence band for the CDF F . Repeat this 1000 times and see how often the confidence band contains the true distribution function. Repeat using data from a Cauchy distribution.
 - If you are not familiar with confidence band of CDF, read Technical appendix of chapter 8 in course textbook

3. Use old faithful geyser dataset to:
 - a) Estimate the mean duration of eruption and give a standard error for the estimate.
 - b) Estimate a 90 percent confidence interval for the mean duration of eruption.
 - c) Estimate the median duration of eruption and give a standard error for the estimate.
4. Consider the magnitude of earthquakes near Fiji which are assumed iid.
 - a) Estimate the CDF, $F(x)$. Plot estimated CDF.
 - b) Compute and plot a 95 percent confidence band for F .
 - c) Approximate 95 percent confidence interval for $(4.9)-F(4.3)$. (Try all 3 types of confidence intervals: pivotal, normal and percentile)
5. Let $X_1, \dots, X_n \sim \text{Normal}(\mu, 1)$. Let $\theta = e^\mu$ and $\hat{\theta} = e^{\bar{X}}$ be the MLE. Create a dataset ($\mu = 10$) consisting of $n=100$ observations.
 - a) Use the bootstrap to get the standard error and 95 percent confidence interval for θ .
 - b) Plot a histogram of the bootstrap replications for the nonparametric bootstrap. These are estimates of $\hat{\theta}$ distribution. Compare this to the true sampling distribution of $\hat{\theta}$.
6. Let $X_1, \dots, X_n \sim \text{Normal}(\mu, 1)$.
 - a) Simulate a dataset (using $\mu=10$) consisting of $n=100$ observations.
 - b) Take $f(\mu)=1$ and find the posterior density. Plot the density.
 - c) Simulate 1000 draws from posterior. Plot a histogram of the simulated value and compare the histogram to the answer in b.
 - d) Let $\theta=e^\mu$. Find the posterior density for θ analytically and by simulation.
 - e) Find a 95% pivotal interval for θ .
 - f) Find a 95% percentile interval for θ .
7. Generate 100 samples from a beta distribution ($\alpha=2$ and $\beta=5$).
 - a) Calculate & plot empirical distribution function.
 - b) Find the plug-in estimator for:
 - a. Mean
 - b. Variance (both reasonable estimators)
 - c. Skewness