Homework

Due: 24/8/2020 11:59PM

- As mentioned in class, computing the mean $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$ and variance $\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2}$ over the data set $D = \{x_1, \ldots, x_n\}$ is our intuition towards estimating (fitting) the Gaussian distribution into the data.
- It is your homework to continue showing that computing the mean and variance over the data set is exactly finding the optimal parameter $\theta_{MLE} = \{\mu, \sigma\}$ by MLE.

MLE Recap (1)

- There are three steps involving when applying MLE :
  - **Step 1:** You assume the modeling distribution $f(x;\theta)$.
  - **Step 2:** Define the likelihood function $L(\theta | D = \{x_1, \ldots, x_n\}) = \prod_{i}^{n} f(x_i;\theta)$

    Note that $\prod$ *denotes the product of the following terms*
  - **Step 3:** Solve for $\theta_{MLE} = argmax_{\theta} L(\theta | D)$

MLE Recap (2)

- At step 3, we usually solve for θ_{MLE} by considering the first conditional derivative of the log of the likelihood function
  - **Step 3.1:** We the define the function $L'(\theta \mid D) = \log_e L(\theta \mid D)$
  - **Step 3.2:** We derive $\dfrac{\partial L'(\theta \mid D)}{\partial \theta}$
  - **Step 3.3:** We solve for θ_{MLE} that makes $\dfrac{\partial L'(\theta \mid D)}{\partial \theta} = 0$
- **Remarks:** This recap is more like a cookbook for deriving the θ_{MLE}. However, you should be aware how each step is derived. This will gives you a full understanding of MLE.

Programming Assignment III

Due: 24/8/2020 11:59PM

- **Task:** Let's try to estimate the probability distribution of the length of Mackerels and Salmons using MLE by assuming *Gaussian distribution* for the modeling distributions.

- **Data set:** fish_dataset.csv

  - First column: fish length

  - Second column: class (1=Mackerels; 2=Salmon)

- **Submission:** You are to plot the distributions of the length of Mackerels and Salmons estimated by MLE.

Expected Solution to Programming Assignment III

- Your code must contain the portion that gives the attached distribution plot.

```
x = np.linspace(0, 80, 100)

x_test = 45

plt.plot(x, gaussain_mackarels(x), linewidth=2,
color='g', label="Mackerels")

plt.plot(x, gaussain_salmons(x), linewidth=2,
color='r', label="Salmons")

plt.plot(x_test, gaussain_mackarels(x_test), 'go')
plt.plot(x_test, gaussain_salmons(x), 'ro')
plt.axvline(x_test, -1, 1, linestyle='--',
label="Likelihood of x_test = 45")
plt.legend()
plt.show()
```