Data-Driven Design & Analysis of Structures & Materials (3dasm)

Intructor: Miguel A. Bessa and **TAs**: Martin van der Schelling & Bruno Ribeiro

Brown University

Homework 3

Deliver **a short PDF report** of this assignment containing the answers to the questions listed here. **UPLOAD to CANVAS in the Assignments section (Homework 3) by the due date.**

Due date: September 27, 2023 (until 11:59pm)

Exercise 1

1. Show that the product of two Gaussian probability density functions (pdf's) for the same random variable (rv) z is:

$$\mathcal{N}(z|\mu_1, \sigma_1^2) \cdot \mathcal{N}(z|\mu_2, \sigma_2^2) = C \cdot \mathcal{N}(z|\mu, \sigma^2)$$

where

$$\begin{aligned}
\sigma^2 &= \frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}, \\
\mu &= \sigma^2 \left(\frac{\mu_1}{\sigma_1^2} + \frac{\mu_2}{\sigma_2^2} \right), \\
C &= \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} \exp\left[-\frac{1}{2(\sigma_1^2 + \sigma_2^2)} (\mu_1 - \mu_2)^2 \right].
\end{aligned}$$

Exercise 2

2. Returning to our favorite problem of the car stopping distance for a constant velocity of $x = 75$ m/s, and considering the following model choices:

   - Observation distribution:

$$p(y|z) = \mathcal{N}\left(y|\mu_{y|z} = wz + b, \sigma_{y|z}^2\right) = \frac{1}{C_{y|z}} \exp\left[-\frac{1}{2\sigma_{y|z}^2} (y - \mu_{y|z})^2 \right]$$

     where $C_{y|z} = \sqrt{2\pi\sigma_{y|z}^2}$ is the **normalization constant** of the Gaussian pdf, and where $\mu_{y|z} = wz + b$, with w, b and $\sigma_{y|z}^2$ being constants.

   - and prior distribution: $p(z) = \frac{1}{C_z}$

     where $C_z = z_{max} - z_{min}$ is the normalization constant of the Uniform pdf, i.e. the value that guarantees that $p(z)$ integrates to one. Consider the domain for z to be $-\infty$ to ∞.

   In Lecture 6 we determined that the Posterior Predictive Distribution (PPD) for this case is:

$$p(y|\mathcal{D}_y) = \frac{1}{|w|} \frac{1}{\sqrt{2\pi\left(\sigma^2 + \frac{\sigma_{y|z}^2}{w^2}\right)}} \exp\left[-\frac{\left(\mu - \frac{y-b}{w}\right)^2}{2\left(\sigma^2 + \frac{\sigma_{y|z}^2}{w^2}\right)} \right]$$

Rewrite the PPD to show that it becomes:

$$p(y|\mathcal{D}_y) = \mathcal{N}\left(y|b + \mu w, w^2\sigma^2 + \sigma^2_{y|z}\right)$$

a normalized univariate Gaussian!

Exercise 3

3. Consider again the same car stopping distance problem, but now defining a model with a different prior distribution:

   - Observation distribution:

   $$p(y|z) = \mathcal{N}\left(y|\mu_{y|z} = wz + b, \sigma^2_{y|z}\right) = \frac{1}{C_{y|z}} \exp\left[-\frac{1}{2\sigma^2_{y|z}}(y - \mu_{y|z})^2\right]$$

   where $C_{y|z} = \sqrt{2\pi\sigma^2_{y|z}}$ is the **normalization constant** of the Gaussian pdf, and where $\mu_{y|z} = wz + b$, with w, b and $\sigma^2_{y|z}$ being constants.

   - and prior distribution: $p(z) = \mathcal{N}\left(z|\overset{<}{\mu}_z = 3, \overset{<}{\sigma}^2_z = 2^2\right)$

   Follow the Bayesian approach and determine the PPD using this model.

Exercise 4[1]

4. Plot[2] the PPD obtained in Exercise 3 and in Exercise 4 together on the same figure with different line thicknesses and colors such that you can visualize the two PPDs at the same time. Create two plots considering a different number of samples, i.e. for y_i with $i = 1, ..., N$ where:

   4.1. $N = 5$ samples.

   4.2. $N = 300$ samples.

Exercise 5

5. Calculate point estimates for different problems:

   5.1. Using the MLE point estimate, predict the PPD for the car stopping distance problem with the Gaussian observation distribution and the Uniform prior considered in Exercise 2.

   5.2. Using the MAP estimate, predict the PPD for the car stopping distance problem with the Gaussian observation distribution and the Gaussian prior considered in Exercise 3.

   5.3. Create a plot[3] of the two PPD's and compare them with the PPD's obtained in Exercise 2 and Exercise 3 (in other words, the PPD's you plotted in Exercise 4).

   Note: create these plots of the PPD's such that the abscissa (horizontal) axis is the y rv and the ordinate (vertical axis) is the probability density.

   As in Exercise 4, make these plots considering a different number of samples y_i:

   5.3.1. $N = 5$ samples.

   5.3.2. $N = 300$ samples.

[1] Coding exercise
[2] Code is available in Lecture 6
[3] Coding exercise