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

Lecture 7

Miguel A. Bessa | **miguel_bessa@brown.edu** | Associate Professor

Outline for today

- Understanding the Posterior Predictive Distribution (PPD)
    - Similar example to Lecture 6 but now with a Gaussian prior.

Reading material: This notebook

In this Lecture...

Summary of the model

1. The **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.

1. but now assuming a different **prior distribution**: $p(z) = \mathcal{N}\left(z|\overset{<}{\mu}_z = 3, \overset{<}{\sigma}{}_z^2 = 2^2\right)$

As in Lecture 6, we start by using Bayes' rule applied to data to determine the posterior:

$$p(z|y = \mathcal{D}_y) = \frac{p(y = \mathcal{D}_y|z)p(z)}{p(y = \mathcal{D}_y)}$$

The likelihood is the same as in Lecture 6:

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

where $\mu = \frac{w^2\sigma^2}{\sigma^2_{y|z}} \sum_{i=1}^N \mu_i = \frac{\sum_{i=1}^N y_i}{wN} - \frac{b}{w}$

$\sigma^2 = \frac{\sigma^2_{y|z}}{w^2 N}$, and

$C = \frac{1}{2\pi^{(N-1)/2}} \sqrt{\frac{\sigma^2}{\left(\frac{\sigma^2_{y|z}}{w^2}\right)^N}}$

But now the marginal likelihood is different from Lecture 6 because we have a different prior:

$$p(y = \mathcal{D}_y) = \frac{C \cdot C_M}{|w|^N}$$

where $C_M = \frac{1}{\sqrt{2\pi\left(\sigma^2 + \overset{<}{\sigma}_z^2\right)}} \exp\left[-\frac{1}{2\left(\sigma^2 + \overset{<}{\sigma}_z^2\right)}\left(\mu - \overset{<}{\mu}_z\right)^2\right]$.

(Algebra to get this result is in the notes below.)

Therefore, the posterior will also be different:

$$
\begin{aligned}
p(z|y=\mathcal{D}_y) &= \frac{p(y=\mathcal{D}_y|z)p(z)}{p(y=\mathcal{D}_y)} && (4)\\
&= \frac{|w|^N}{C \cdot C_M} \cdot \frac{1}{|w|^N} C \cdot \mathcal{N}(z|\mu,\sigma^2) \cdot \mathcal{N}\left(z|\overset{<}{\mu}_z, \overset{<}{\sigma}{}_z^2\right) && (5)\\
&= \mathcal{N}\left(z|\overset{>}{\mu}_z, \overset{>}{\sigma}{}_z^2\right) && (6)
\end{aligned}
$$

where $\overset{>}{\mu}_z = \frac{1}{\frac{1}{\sigma^2}+\frac{1}{\overset{<}{\sigma}{}_z^2}} \left(\frac{\mu}{\sigma^2} + \frac{\overset{<}{\mu}_z}{\overset{<}{\sigma}{}_z^2}\right)$

and $\overset{>}{\sigma}{}_z^2 = \frac{1}{\frac{1}{\sigma^2}+\frac{1}{\overset{<}{\sigma}{}_z^2}}$

are the parameters of the posterior distribution, symbolized by the superscript $\overset{>}{(\cdot)}$.

Reflection on the differences between the posterior we obtain for the two different priors we considered.

- When using the noninformative Uniform prior $p(z) = \frac{1}{C_z}$ (Lecture 6):

$$p(z|y = \mathcal{D}_y) = \mathcal{N}(z|\mu, \sigma^2) \tag{7}$$

- When using a Gaussian prior $p(z) = \mathcal{N}\left(z|\overset{<}{\mu}_z, \overset{<}{\sigma}_z^2\right)$ (this Lecture):

$$p(z|y = \mathcal{D}_y) = \mathcal{N}\left(z|\overset{>}{\mu}_z, \overset{>}{\sigma}_z^2\right) = \mathcal{N}\left(z\left|\frac{1}{\frac{1}{\sigma^2} + \frac{1}{\overset{<}{\sigma}_z^2}}\left(\frac{\mu}{\sigma^2} + \frac{\overset{<}{\mu}_z}{\overset{<}{\sigma}_z^2}\right), \frac{1}{\frac{1}{\sigma^2} + \frac{1}{\overset{<}{\sigma}_z^2}}\right.\right) \tag{8}$$

The posterior is still a Gaussian but its mean and variance have been updated by the influence of the prior!

Finally, the goal of calculating the posterior is to use it to determine the Posterior Predictive Distribution (PPD) :

$$p(y^*|\mathcal{D}_y) = \int \underbrace{p(y^*|z)}_{\substack{\text{observation}\\\text{distribution}}} \overbrace{p(z|y = \mathcal{D}_y)}^{\text{posterior}} dz$$

where y^* highlights that we can make a prediction for any y value (just to highlight that it has nothing to do with the training data).

Considering the terms we found before, we get:

$$p(y^*|\mathcal{D}_y) = \int \underbrace{\frac{1}{|w|} \frac{1}{\sqrt{2\pi \left(\frac{\sigma_{y|z}}{w}\right)^2}} \exp\left\{ -\frac{1}{2\left(\frac{\sigma_{y|z}}{w}\right)^2} \left[z - \left(\frac{y^* - b}{w} \right) \right]^2 \right\}}_{\substack{\text{observation} \\ \text{distribution}}} \overbrace{\mathcal{N}\left(z|\overset{>}{\mu}_z, \overset{>}{\sigma}{}_z^2\right)}^{\text{posterior}} dz \tag{9}$$

The calculation of this integral is similar to what we did in Lecture 6! The difference is that the posterior has a different mean and variance (indicated with the superscript) that originated from the choice of different prior!

So, we can fast forward to the result we obtained before! We just need to replace the symbols μ_z by $\overset{>}{\mu}_z$, and σ_z^2 for $\overset{>}{\sigma}_z^2$:

$$p(y^*|\mathcal{D}_y) = \frac{\tilde{C}}{|w|}$$

where

$$\tilde{C} = \frac{1}{\sqrt{2\pi\left(\overset{>}{\sigma}_z^2 + \frac{\sigma_{y|z}^2}{w^2}\right)}} \exp\left[-\frac{\left(\overset{>}{\mu}_z - \frac{y^*-b}{w}\right)^2}{2\left(\overset{>}{\sigma}_z^2 + \frac{\sigma_{y|z}^2}{w^2}\right)}\right]$$

is the same constant as C^* in Lecture 6, but replacing μ_z by $\overset{>}{\mu}_z$, and σ_z^2 for $\overset{>}{\sigma}_z^2$.

After a bit of algebra, we get to the following expression for the PPD:

$$
\begin{aligned}
p(y^*|\mathcal{D}_y) &= \frac{1}{\sqrt{2\pi\left(\sigma_{y|z}^2 + w^2\vec{\sigma}_z^2\right)}}\exp\left\{-\frac{1}{2\left(\sigma_{y|z}^2 + w^2\vec{\sigma}_z^2\right)}\left[y^* - \left(w\vec{\mu}_z + b\right)\right]^2\right\} && (10)\\
&= \mathcal{N}\left(y^* \middle| w\vec{\mu}_z + b, \sigma_{y|z}^2 + w^2\vec{\sigma}_z^2\right) && (11)
\end{aligned}
$$

where all of the terms have been defined before (for convenience, see them in the next cell as notes).

In order to see the explicit dependence on the observed data $\mathcal{D}_y$, we can also rewrite the PPD as:

$$p(y^*|\mathcal{D}_y) = \mathcal{N}\left(y^* \,|\mu^*, \sigma^*\right)$$

where

$\mu^* = \frac{1}{1+\frac{\sigma_{y|z}^2}{w^2 N \overset{<}{\sigma}_z^2}} \left[\frac{\sum_{i=1}^N y_i}{N} + \frac{\sigma_{y|z}^2}{w^2 N \overset{<}{\sigma}_z^2} \left(w \overset{<}{\mu}_z + b \right) \right]$ (**Note** that y_i is the training data point i, not the new predicted value y^*)

$$\left(\sigma^*\right)^2 = \sigma_{y|z}^2 + w^2 \overset{>}{\sigma}_z^2 = \sigma_{y|z}^2 + \frac{\overset{<}{\sigma}_z^2 \frac{\sigma_{y|z}^2}{N}}{\overset{<}{\sigma}_z^2 + \frac{\sigma_{y|z}^2}{w^2 N}}$$

- What happens when $N \to \infty$?

When $N \to \infty$ the mean and variance of the PPD become:

$$\mu^* = \frac{\sum_{i=1}^N y_i}{N} \equiv \text{Empirical mean}$$

$$(\sigma^*)^2 = \sigma^2_{y|z} \equiv \text{Variance assumed by us for the observation distribution}$$

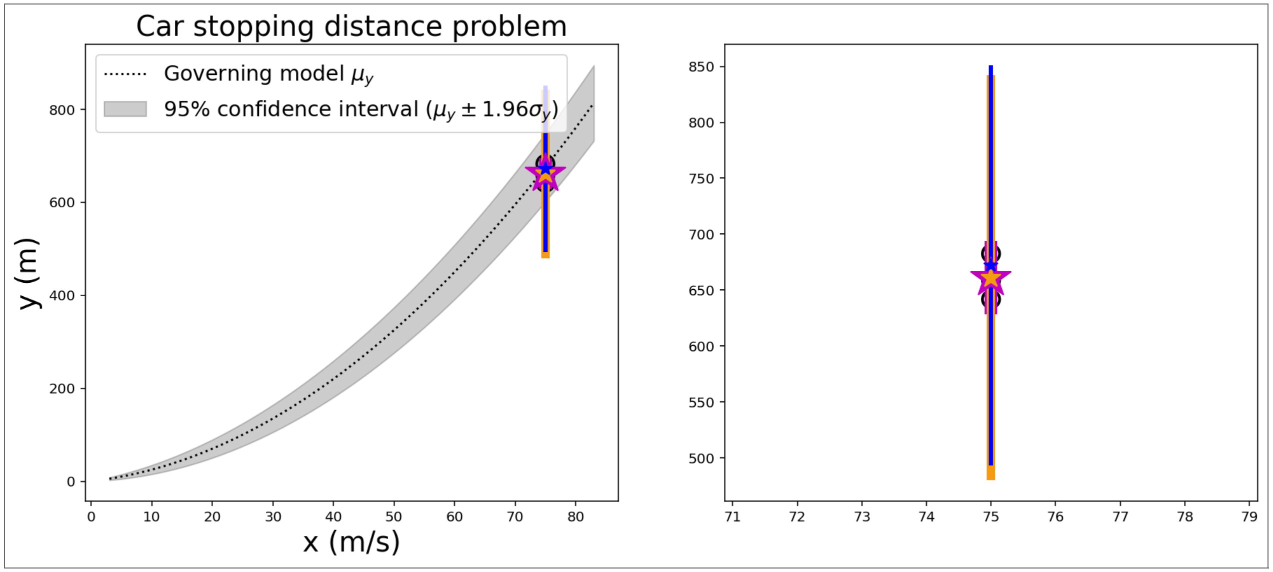
So, in this limit the PPD is simply:

$$p(y^*|\mathcal{D}_y) = \mathcal{N}\left(y^* \middle| \frac{\sum_{i=1}^N y_i}{N}, \sigma^2_{y|z}\right) \quad \text{when } N \to \infty$$

- This means that in the limit of $N \to \infty$ we have exactly the same result obtained when we used the noninformative Uniform prior! Were you expecting this? Let's debate!

```
In [5]: PPD_comparison(N_samples=3) # Plot data and the two PPD's considering different priors
```

```
Ground truth                    : mean[y] = 675    & std[y] = 37.5
Empirical values (purple)       : mean[y] = 661.04 & std[y] = 16.71
PPD with Uniform Prior (orange): mean[y] = 661.04 & std[y] = 92.38
PPD with Gaussian Prior (blue) : mean[y] = 671.99 & std[y] = 91.37
```



See you next class

Have fun!