Data-driven Design and Analyses of Structures and Materials (3dasm)

Lecture 5

Miguel A. Bessa | **M.A.Bessa@tudelft.nl** | Associate Professor

OPTION 1. Run this notebook **locally in your computer**:

1. Confirm that you have the 3dasm conda environment (see Lecture 1).
2. Go to the 3dasm_course folder in your computer and pull the last updates of the **repository**:

```
git pull
```

3. Open command window and load jupyter notebook (it will open in your internet browser):

```
conda activate 3dasm
jupyter notebook
```

4. Open notebook of this Lecture.

OPTION 2. Use **Google's Colab** (no installation required, but times out if idle):

1. go to **https://colab.research.google.com**
2. login
3. File > Open notebook
4. click on Github (no need to login or authorize anything)
5. paste the git link: **https://github.com/bessagroup/3dasm_course**
6. click search and then click on the notebook for this Lecture.

Outline for today

- Bayesian inference for one hidden rv
    - Prior
    - Likelihood
    - Marginal likelihood
    - Posterior
    - Gaussian pdf's product

Reading material: This notebook + Chapter 3

Recall the "slightly more complicated" car stopping distance problem (with two rv's)

We defined the governing model with two rv's z_1 and z_2 as:

$y = z_1 \cdot x + z_2 \cdot x^2$

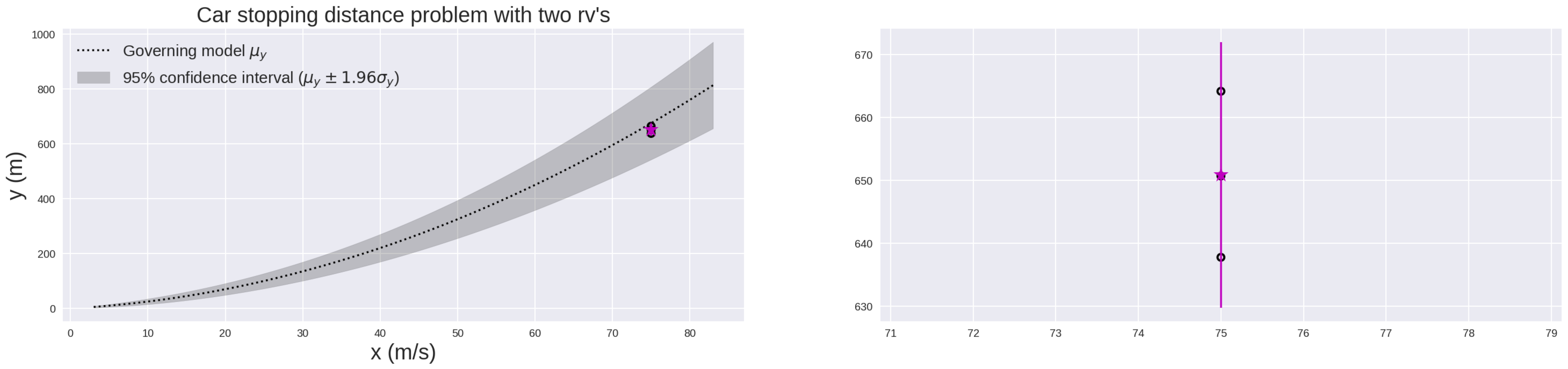
- y is the **output**: the car stopping distance (in meters)
- z_1 is an rv representing the driver's reaction time (in seconds)
- z_2 is another rv that depends on the coefficient of friction, the inclination of the road, the weather, etc. (in $\mathrm{m}^{-1}\mathrm{s}^{-2}$).
- x is the **input**: constant car velocity (in m/s).

where we knew the "true" distributions of the rv's: $z_1 \sim \mathrm{N}(\mu_{z_1} = 1.5, \sigma^2_{z_1} = 0.5^2)$, and $z_2 \sim \mathrm{N}(\mu_{z_2} = 0.1, \sigma^2_{z_2} = 0.01^2)$.

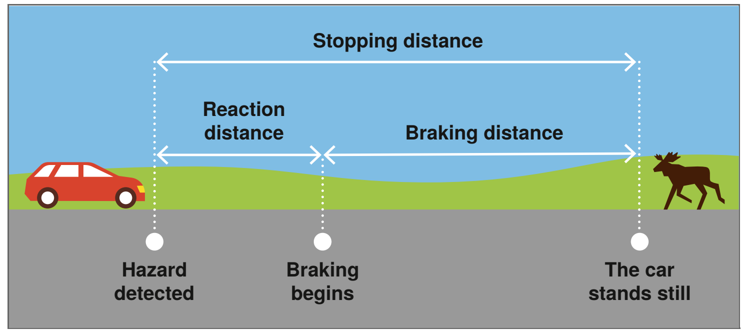
In [4]:

```
            # vvvvvvvvvvv this is just a trick so that we can run this cell multiple times vvvvvvvvvvv
fig_car_new, ax_car_new = plt.subplots(1,2); plt.close() # create figure and close it
if fig_car_new.get_axes():
    del ax_car_new; del fig_car_new # delete figure and axes if they exist
    fig_car_new, ax_car_new = plt.subplots(1,2) # create them again
# ^^^^^^^^^^^ end of the trick ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
N_samples = 3 # CHANGE THIS NUMBER AND RE-RUN THE CELL
x = 75; empirical_y = samples_y_with_2rvs(N_samples, x); # Empirical measurements of N_samples at x=75
empirical_mu_y = np.mean(empirical_y); empirical_sigma_y = np.std(empirical_y); # empirical mean and std
car_fig_2rvs(ax_car_new[0]) # a function I created to include the background plot of the governing model
for i in range(2): # create two plots (one is zooming in on the error bar)
    ax_car_new[i].errorbar(x , empirical_mu_y,yerr=1.96*empirical_sigma_y, fmt='m*', markersize=15);
    ax_car_new[i].scatter(x*np.ones_like(empirical_y),empirical_y, s=40,
                          facecolors='none', edgecolors='k', linewidths=2.0)
print("Empirical mean[y] is",empirical_mu_y, "(real mean[y]=675)")
print("Empirical std[y] is",empirical_sigma_y,"(real std[y]=67.6)")
fig_car_new.set_size_inches(25, 5) # scale figure to be wider (since there are 2 subplots)
```

```
Empirical mean[y] is 650.874924199532 (real mean[y]=675)
Empirical std[y] is 10.765521210487918 (real std[y]=67.6)
```



Car stopping distance problem with 2 rv's but only 1 rv being unknown



Today we will finally do some predictions!
Recall the Homework of Lecture 4, and consider the car stopping distance problem for constant velocity $x = 75$ m/s and for which **it is known** that $z_2 \sim \mathcal{N}(z_2 | \mu_{z_2} = 0.1, \sigma_{z_2}^2 = 0.01^2)$.

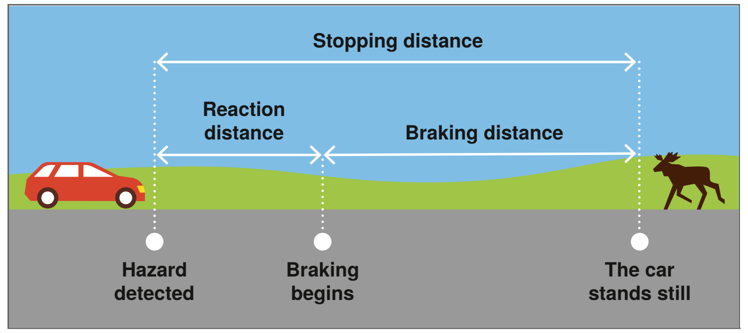
The only information that we do not know is the driver's reaction time z (here we call it z, instead of z_1 as in Lecture 4, because this is the only hidden variable so we can **simplify the notation**).

- Can we predict $p(y)$ without knowing $p(z)$?

Yes!! If we use Bayes' rule!

Recall the Homework of Lecture 4

From last lecture's Homework, you demonstrated that the conditional pdf of the stopping distance given the reaction time z (for convenience we write here z instead of z_1) is



$$p(y\,|\,z) = \mathrm{N}\left(y\,|\,\mu_{y|z} = wz + b, \sigma^2_{y|z} = s^2\right)$$

where w, b and s are all constants that you determined to be:

$w = x = 75$

$b = x^2\mu_{z_2} = 75^2 \cdot 0.1 = 562.5$

$s^2 = (x^2\sigma_{z_2})^2 = (75^2 \cdot 0.01)^2 = 56.25^2$

because we are considering that the car is going at constant velocity $x = 75$ m/s and that we know

$z_2 = \mathrm{N}(z_2\,|\,\mu_{z_2} = 0.1, \sigma^2_{z_2} = 0.01^2)$.

Understanding the Bayes' rule

$$p(z \mid y) = \frac{p(y \mid z)p(z)}{p(y)}$$

- $p(z)$ is the **prior distribution**
- $p(y \mid z)$ is the **observation distribution** (conditional pdf)
- $p(y)$ is the **marginal distribution**
- $p(z \mid y)$ is the **posterior distribution**

Understanding the Bayes' rule

Let's start by understanding the usefulness of Bayes' rule by calculating the posterior $p(z|y)$ for the car stopping distance problem (Homework of Lecture 4).

As we mentioned, for our problem we know the **observation distribution**:

$$p(y|z) = \mathrm{N}\left(y|\mu_{y|z} = wz + b, \sigma^2_{y|z}\right)$$

where $\sigma_{y|z} = \mathrm{const}$, as well as w and b.

but we **don't know** the prior $p(z)$.

Prior: our beliefs about the problem

If we have absolutely no clue about what the distribution of the hidden rv z is, then we can use a **Uniform distribution** (a.k.a. uninformative prior).
This distribution assigns equal probability to any value of z within an interval $z \in (z_{min}, z_{max})$.

$$p(z) = \frac{1}{C_z}$$

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.
For the time being, we will not assume any particular values for z_{max} and z_{min}. So, we will consider the completely uninformative prior: $z_{max} \to \infty$ and $z_{min} \to -\infty$. If we had some information, we could consider some values for these bounds (e.g. $z_{min} = 0$ seconds would be the limit of the fastest reaction time that is humanly possible, and $z_{max} = 3$ seconds would be the slowest reaction time of a human being).

Summary of our Model

1. The **observation distribution**:

$$p(y|z) = \mathrm{N}\left(y \,|\, \mu_{y|z} = wz + b, \sigma^2_{y|z}\right)$$

$$= \frac{1}{C_{y|z}} \exp\left[-\frac{1}{2\sigma^2_{y|z}} (y - \mu_{y|z})^2 \right]$$

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, as previously mentioned.

1. and the **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.

Posterior from Bayes' rule

Since we have defined the **observation distribution** and the **prior distribution**, we can now compute the posterior distribution from Bayes' rule.

But this requires a bit of algebra... Let's do it!

First, in order to apply Bayes' rule $p(z|y) = \frac{p(y|z)p(z)}{p(y)}$ we need to calculate $p(y)$.

$p(y)$ is obtained by marginalizing the joint distribution wrt z:
$p(y) = \int p(y|z)p(z)dz$
which implies an integration over z. So, let's rewrite $p(y|z)$ so that the integration becomes easier.

$$
\begin{aligned}
p(y \mid z) &= \mathrm{N}\left(y \mid \mu_{y|z} = wz + b, \sigma^2_{y|z}\right) \\
&= \frac{1}{C_{y|z}} \exp\left[-\frac{1}{2\sigma^2_{y|z}} (y - (wz + b))^2 \right] \\
&= \frac{1}{C_{y|z}} \exp\left\{ -\frac{1}{2\left(\frac{\sigma_{y|z}}{w}\right)^2} \left[z - \left(\frac{y-b}{w}\right) \right]^2 \right\} \\
&= \frac{1}{|w|} \frac{1}{\sqrt{2\pi\left(\frac{\sigma_{y|z}}{w}\right)^2}} \exp\left\{ -\frac{1}{2\left(\frac{\sigma_{y|z}}{w}\right)^2} \left[z - \left(\frac{y-b}{w}\right) \right]^2 \right\}
\end{aligned}
$$

Note: This Gaussian pdf $\mathrm{N}\left(z \mid \frac{y-b}{w}, \left(\frac{\sigma_{y|z}}{w}\right)^2\right)$ is unnormalized when written wrt z (due to $\frac{1}{|w|}$).

We can now calculate the marginal distribution $p(y)$:

$$p(y) = \int p(y\,|\,z)p(z)dz$$

$$= \int \frac{1}{|w|} \frac{1}{\sqrt{2\pi \left(\frac{\sigma_{y|z}}{w}\right)^2}} \exp\left\{ -\frac{1}{2\left(\frac{\sigma_{y|z}}{w}\right)^2} \left[z - \left(\frac{y-b}{w}\right) \right]^2 \right\} \frac{1}{C_z} dz$$

We can rewrite this expression as,

$$p(y) = \frac{1}{|w| \cdot C_z} \int \frac{1}{\sqrt{2\pi \left(\frac{\sigma_{y|z}}{w}\right)^2}} \exp\left\{ -\frac{1}{2\left(\frac{\sigma_{y|z}}{w}\right)^2} \left[z - \left(\frac{y-b}{w}\right) \right]^2 \right\} dz$$

What is the result for the blue term?

From where we conclude that the marginal distribution is:

$$p(y) = \frac{1}{|w|C_z}$$

So, now we can determine the posterior:

$$p(z\,|\,y) = \frac{p(y\,|\,z)p(z)}{p(y)}$$

$$= |w|C_z \cdot \frac{1}{|w|}\frac{1}{\sqrt{2\pi\left(\frac{\sigma_{y|z}}{w}\right)^2}}\exp\left\{-\frac{1}{2\left(\frac{\sigma_{y|z}}{w}\right)^2}\left[z-\left(\frac{y-b}{w}\right)\right]^2\right\}\cdot\frac{1}{C_z}$$

$$= \frac{1}{\sqrt{2\pi\left(\frac{\sigma_{y|z}}{w}\right)^2}}\exp\left\{-\frac{1}{2\left(\frac{\sigma_{y|z}}{w}\right)^2}\left[z-\left(\frac{y-b}{w}\right)\right]^2\right\}$$

which is a **normalized** Gaussian pdf in z: $\mathrm{N}\left(z\,|\,\frac{y-b}{w}, \left(\frac{\sigma_{y|z}}{w}\right)^2\right)$

- **This is what the Bayes' rule does!** Computes the posterior $p(z\,|\,y)$ from $p(y\,|\,z)$ and $p(z)$.

Why should we care about the Bayes' rule?

There are a few reasons:

1. As we will see, models are usually (always?) wrong.
1. But our beliefs may be a bit closer to reality! Bayes' rule enables us to get better models if our beliefs are reasonable!
1. We don't observe distributions. We observe **DATA**. Bayes' rule is a very powerful way to predict the distribution of our quantity of interest (here: y) from data!

Bayes' rule applied to observed data

Previously, we already introduced Bayes' rule when applied to observed data D_y.

$$p(z \mid y = D_y) = \frac{p(y = D_y \mid z)p(z)}{p(y = D_y)} = \frac{p(y = D_y, z)}{p(y = D_y)}$$

- $p(z)$ is the **prior** distribution
- $p(y = D_y \mid z)$ is the **likelihood** function
- $p(y = D_y, z)$ is the **joint likelihood** (product of likelihood function with prior distribution)
- $p(y = D_y)$ is the **marginal likelihood**
- $p(z \mid y = D_y)$ is the **posterior**

We can write Bayes' rule as posterior $\propto$ likelihood $\times$ prior , where we are ignoring the denominator $p(y = D_y)$ because it is just a **constant** independent of the hidden variable z.

Bayes' rule applied to observed data

But remember that Bayes' rule is just a way to calculate the posterior:

$$p(z \mid y = D_y) = \frac{p(y = D_y \mid z)p(z)}{p(y = D_y)}$$

Usually, what we really want is to be able to predict the distribution of the quantity of interest (here: y) after observing some data D_y:

$$p(y \mid y = D_y) = \int p(y \mid z)p(z \mid y = D_y)dz$$

which is often written in simpler notation: $p(y \mid D_y) = \int p(y \mid z)p(z \mid D_y)dz$

Bayesian inference for car stopping distance problem

Now we will solve the first Bayesian ML problem from some given data $y = \mathrm{D}_y$:

y_i (m)
601.5
705.9
693.8
...
711.3

where the data D_y could be a Pandas dataframe with N data points (N rows).

- **Very Important Question (VIQ)**: Can we calculate the **likelihood** function from this data?

Likelihood for car stopping distance problem

Of course! As we saw a few cells ago, the **likelihood** is obtained by evaluating the **observation distribution** at the data D_y. Noting that each observation in D_y is independent of each other, then:

$$p(y = \mathrm{D}_y | z) = \prod_{i=1}^{N} p(y = y_i | z) = p(y = y_1 | z) p(y = y_2 | z) \cdots p(y = y_N | z)$$

which gives the **probability density** of observing that data if using our observation distribution (part of our model!).

Calculating the likelihood

Let's calculate it:

$$p(y = \mathrm{D}_y \,|\, z) = \prod_{i=1}^{N} p(y = y_i \,|\, z)$$

$$= \prod_{i=1}^{N} \frac{1}{C_{y|z}} \exp\left\{ -\frac{1}{2\left(\frac{\sigma_{y|z}}{w}\right)^2} \left[z - \left(\frac{y_i - b}{w} \right) \right]^2 \right\}$$

This seems a bit daunting... I know. Do not dispair yet!

Product of Gaussian pdf's of the same rv z

It can be shown that the product of N univariate Gaussian pdf's of the same rv z is:

$$\prod_{i=1}^{N} \mathrm{N}(z \mid \mu_i, \sigma_i^2) = C \cdot \mathrm{N}(z \mid \mu, \sigma^2)$$

with mean: $\mu = \sigma^2 \left(\sum_{i=1}^{N} \frac{\mu_i}{\sigma_i^2} \right)$

variance: $\sigma^2 = \frac{1}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}}$

and normalization constant: $C = \frac{1}{(2\pi)^{(N-1)/2}} \sqrt{\frac{\sigma^2}{\prod_{i=1}^{n} \sigma_i^2}} \exp\left[-\frac{1}{2} \left(\sum_{i=1}^{N} \frac{\mu_i^2}{\sigma_i^2} - \frac{\mu^2}{\sigma^2} \right) \right]$

Curiosity: the normalization constant C is itself a Gaussian! You can see it more clearly if you consider $N = 2$

HOMEWORK

Show that the product of two Gaussian pdf's for the same rv z is:

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

$$\sigma^2 = \frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}$$

$$\mu = \sigma^2 \left(\frac{\mu_1}{\sigma_1^2} + \frac{\mu_2}{\sigma_2^2} \right)$$

$$C = \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} \exp\left[-\frac{1}{2(\sigma_1^2 + \sigma_2^2)} (\mu_1 - \mu_2)^2 \right]$$

Back to calculating the likelihood

$$p(y = \mathrm{D}_y | z) = \prod_{i=1}^{N} p(y = y_i | z)$$

$$= \prod_{i=1}^{N} \frac{1}{|w|} \frac{1}{\sqrt{2\pi\left(\frac{\sigma_{y|z}}{w}\right)^2}} \exp\left\{ -\frac{1}{2\left(\frac{\sigma_{y|z}}{w}\right)^2} \left[z - \left(\frac{y_i - b}{w}\right) \right]^2 \right\}$$

$$= \frac{1}{|w|^N} \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\left(\frac{\sigma_{y|z}}{w}\right)^2}} \exp\left\{ -\frac{1}{2\left(\frac{\sigma_{y|z}}{w}\right)^2} \left[z - \left(\frac{y_i - b}{w}\right) \right]^2 \right\}$$

So, using the result of a product of N Gaussian pdf's to calculate the likelihood, and noting that $\sigma_i = \frac{\sigma_{y|z}}{w}$ and $\mu_i = \frac{y_i - b}{w}$ we get:

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

where

$$\mu = \frac{\sigma^2}{\sigma_i^2} \sum_{i=1}^N \mu_i = \frac{w^2\sigma^2}{\sigma_{y|z}^2} \sum_{i=1}^N \mu_i$$

$$\sigma^2 = \frac{1}{\sum_{i=1}^N \frac{1}{\sigma_i^2}} = \frac{1}{\sum_{i=1}^N \frac{w^2 N}{\sigma_{y|z}^2}} = \frac{\sigma_{y|z}^2}{\sum_{i=1}^N w^2 N}$$

$$C = \frac{1}{(2\pi)^{(N-1)/2}} \sqrt{\frac{\sigma^2}{\left(\frac{\sigma_{y|z}^2}{w^2}\right)^N}} \exp\left[-\frac{1}{2}\left(\frac{w^2}{\sigma_{y|z}^2} \sum_{i=1}^N \mu_i - \frac{\mu^2}{\sigma^2} \right) \right]$$

Calculating the marginal likelihood

$$
\begin{aligned}
p(y = \mathrm{D}_y) &= \int p(y = \mathrm{D}_y | z) p(z) dz \\
&= \int \frac{1}{|w|^N} C \cdot \mathrm{N}(z | \mu, \sigma^2) \cdot \frac{1}{C_z} dz \\
&= \frac{C}{|w|^N C_z} \int \mathrm{N}(z | \mu, \sigma^2) dz = \frac{C}{|w|^N C_z}
\end{aligned}
$$

We can now calculate the posterior:

$$
\begin{aligned}
p(z | y = \mathrm{D}_y) &= \frac{p(y = \mathrm{D}_y | z) p(z)}{p(y = \mathrm{D}_y)} \\
&= \frac{1}{p(y = \mathrm{D}_y)} \cdot \frac{1}{|w|^N} C \cdot \mathrm{N}(z | \mu, \sigma^2) \cdot \frac{1}{C_z} \\
&= \mathrm{N}(z | \mu, \sigma^2)
\end{aligned}
$$

Calculating the Posterior Predictive Distribution (PPD)

Having found the posterior, we can determine the PPD:

$$p(y \mid \mathrm{D}_y) = \int p(y \mid z) p(z \mid \mathrm{D}_y) dz$$

To calculate this, we will have to use the identity for a product of two Gaussians.

$$p(y \mid \mathrm{D}_y) = \int \frac{1}{|w|} \mathrm{N}\left(z \mid \frac{y-b}{w}, \left(\frac{\sigma_{y|z}}{w}\right)^2\right) \mathrm{N}(z \mid \mu, \sigma^2) dz$$

$$= \int \frac{1}{|w|} C^* \mathrm{N}\left(z \mid \mu^*, \left(\sigma^*\right)^2\right) dz$$

Next class

In the next class we will finish this example, by solving this integral to determine the PPD $p(y \mid \mathrm{D}_y)$.

See you next class

Have fun!