Se empieza creando el módulo "paths" en donde quedará registrado todos las rutas necesarias para la tarea.

- Se crea la ruta PATH_DATA y PATH_RAW que serán las rutas a las carpetas 'data' y 'raw' respectivamente
- Se crea un dicc. dict_path_raw, que al entregarse de llave 'wNN', entregará la ruta de la carpeta 'wNN' en 'raw'
- Se crean a su vez los dicc.'s dict_csv_mc_a y dict_csv_mc_f, que entrega la ruta para los .csv's 'metrocuadrado_all_wNN' y 'metrocuadrado_furnished_wNN' respectivamente, al entregarle de llave 'wNN'.

Luego se procede a crear un único DataFrame que contenga la informa∅ de los datasets proporcionados. Se realiza como sigue:

- Se crean dos listas de DataFrames, uno para 'all' y otro para 'furnished'.
- Luego se concatenan los DataFrames en dos DataFrames únicos. Uno para 'all' y otro para 'furnished'.
- Paso seguido, se crea una nueva columna, 'furnished', que informa si el DataFrame proviene de 'furnished' (marcado con un 1), o si viene de 'all' (marcado con 0).
- Posteriormente se unen los dos DataFrames creados en un único DataFrame, y se añade la columna extra.
- Finalmente, se eliminan los duplicados.

P3.1

Pdq $p(\vec{w}|\vec{y}, X, \alpha, \beta) \sim N(\vec{m}_N, S_N)$

$p(\vec{w}|\vec{y}, X, \alpha, \beta)$

$$\propto \underbrace{p(\vec{y}|\vec{w}, X, \alpha, \beta)}_{1} \underbrace{p(\vec{w}|X, \alpha, \beta)}_{2} \quad / \text{Bayes}$$

Desarrollemos 1 y 2 por separado:

1: Notemos que $y \perp \alpha$, luego

$$\begin{aligned}
1 &= p(\vec{y}|X, \vec{w}, \beta) \\
&= \prod_{i=1}^{N} p(y_i|\vec{x}_i, \vec{w}, \beta) \quad / \text{iid}, \ y_i \perp \vec{x}_j \ \forall j \neq i \\
&= \prod^{N} N(y_i | \vec{w}^t x_i, \beta^{-1}) \quad / \textstyle\sum_{i=1}^{N} \equiv \sum^{N} \\
&= \prod^{N} (2\pi\beta^{-1})^{-1/2} \exp(-\beta/2 \cdot (y_i - \vec{w}^t \vec{x}_i)^2) \\
&= (2\pi\beta^{-1})^{-N/2} \exp(-\beta/2 \textstyle\sum^{N} (y_i - \vec{w}^t \vec{x}_i)^2) \quad (3)
\end{aligned}$$

2: Notemos que $\vec{w} \perp X, \beta$, luego:

$$\begin{aligned}
2 &= p(\vec{w}|\alpha) \\
&= (2\pi |\alpha^{-1} I|)^{-d/2} \exp(-\tfrac{1}{2} \vec{w}^t (\alpha^{-1} I)^{-1} \vec{w}) \quad / \text{Notando q } \vec{w} \in \mathbb{R}^{[0..d-1]} \\
&= (2\pi \alpha^{-1})^{-d/2} \exp(-\tfrac{\alpha}{2} \vec{w}^t \vec{w}) \quad (4)
\end{aligned}$$

Ocupando los resultados en lo anterior:

$p(\vec{w}|\vec{y}, X, \alpha, \beta)$

$$\propto \exp(-\beta/2 \textstyle\sum^{N} (y_i - \vec{w}^t \vec{x}_i)^2 - \frac{\alpha}{2} \vec{w}^t \vec{w})$$

Si demostramos que

$$\beta \textstyle\sum^{N} (y_i - \vec{w}^t \vec{x}_i)^2 + \alpha \vec{w}^t \vec{w} = (\vec{w} - \vec{m}_N)^t S_N^{-1} (\vec{w} - \vec{m}_N)$$

Podemos concluir que $p(\vec{w}|\vec{y}, X, \alpha, \beta) \sim N(\vec{m}_N, S_N)$.

En efecto:

$$\beta \sum^N (y_i - \vec{w}^t \vec{x}_i)^2 + \alpha \vec{w}^t \vec{w}$$

$$= \alpha \vec{w}^t \vec{w} + \beta \sum^N (y_i^2 - 2 y_i \vec{w}^t \vec{x}_i + (\vec{w}^t \vec{x}_i)^2)$$

$$= \alpha \vec{w}^t \vec{w} + \beta \sum^N y_i^2 - 2\beta \sum^N y_i \vec{w}^t \vec{x}_i + \beta \sum^N (\vec{w}^t \vec{x}_i)^2 \quad (5)$$

Notemos lo sgte:

$$\beta \sum^N (\vec{w}^t \vec{x}_i)^2 = \beta \sum^N (\vec{x}_i^t \vec{w})(\vec{x}_i^t \vec{w})$$

$$= \beta [\vec{x}_i^t \vec{w}]_i^t [\vec{x}_i^t \vec{w}]_i$$ / donde $[\vec{x}_i^t \vec{w}]_i$ denota not. vectorial.

$$= \beta (X\vec{w})^t (X\vec{w}) = \beta \vec{w}^t X^t X \vec{w} \quad (6)$$

$$-2\beta \sum^N y_i \vec{w}^t \vec{x}_i = -2\beta \vec{w}^t [\sum^N \vec{x}_i y_i]$$

$$= -2\beta \vec{w}^t X^t \vec{y} = -\beta \vec{w}^t X^t \vec{y} - \beta \vec{w}^t X^t \vec{y}$$

$$= -\beta \vec{w}^t X^t \vec{y} - \beta \vec{y}^t X \vec{w}$$ / pues si $a \in \mathbb{R} \Rightarrow a^t = a$

$$= -\vec{w}^t S_N^{-1} [\beta S_N X^t \vec{y}] - [\beta \vec{y}^t X S_N^t] S_N^{-1} \vec{w}$$ / pues S_N simétrica.

$$= -\vec{w}^t S_N^{-1} \vec{m}_N - \vec{m}_N^t S_N^{-1} \vec{w} \quad (7)$$

Reemplazando los dos resultados anteriores (6) y (7) en el desarrollo (5):

$$\alpha \vec{w}^t \vec{w} + \beta \sum^N y_i^2 - 2\beta \sum^N y_i \vec{w}^t \vec{x}_i + \beta \sum^N (\vec{w}^t \vec{x}_i)^2$$

$$= \alpha \vec{w}^t \vec{w} + \beta \vec{w}^t X^t X \vec{w} - \vec{w}^t S_N^{-1} \vec{m}_N - \vec{m}_N^t S_N^{-1} \vec{w} + \beta \sum^N y_i^2$$

$$= \vec{w}^t (\alpha I + \beta X^t X) \vec{w} - \vec{w}^t S_N^{-1} \vec{m}_N - \vec{m}_N^t S_N^{-1} \vec{w} + \beta \sum^N y_i^2$$

$$= \vec{w}^t S_N^{-1} \vec{w} - \vec{w}^t S_N^{-1} \vec{m}_N - \vec{m}_N^t S_N^{-1} \vec{w} + \beta \sum^N y_i^2$$

$$= \vec{w}^t S_N^{-1} (\vec{w} - \vec{m}_N) - \vec{m}_N^t S_N^{-1} \vec{w} + \beta \sum^N y_i^2$$

$$= \vec{w}^t S_N^{-1} (\vec{w} - \vec{m}_N) - \vec{m}_N^t S_N^{-1} \vec{w} + \vec{m}_N^t S_N^{-1} \vec{m}_N$$ / Claim: $\beta \sum^N y_i^2 = \vec{m}_N^t S_N^{-1} \vec{m}_N$

$$= \vec{w}^t S_N^{-1} (\vec{w} - \vec{m}_N) - \vec{m}_N^t S_N^{-1} (\vec{w} - \vec{m}_N)$$

$$= (\vec{w}^t - \vec{m}_N^t) S_N^{-1} (\vec{w} - \vec{m}_N) = (\vec{w} - \vec{m}_N)^t S_N^{-1} (\vec{w} - \vec{m}_N)$$

De esta forma, de demostrarse el claim, se concluye que

$$p(\vec{w} \mid \vec{y}, X, \alpha, \beta) \sim N(\vec{m}_N, S_N)$$

qed//