

Aprendizaje de máquinas

Predicciones y regresión no lineal

Felipe Tobar

Facultad de Ciencias Físicas y Matemáticas
Universidad de Chile

Otoño, 2021.

En las clases anteriores se estudió como estimar los parámetros de un modelo. Ahora se verá cómo utilizar estas estimaciones para hacer predicciones. Para esto:

- Se considerará un modelo $p(y|x)$, con parámetro θ , del cual se han obtenido datos denotados mediante $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N$.

En las clases anteriores se estudió como estimar los parámetros de un modelo. Ahora se verá cómo utilizar estas estimaciones para hacer predicciones. Para esto:

- Se considerará un modelo $p(y|x)$, con parámetro θ , del cual se han obtenido datos denotados mediante $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N$.
- Se considerará un modelo de **variable latente**: las observaciones consisten en una variable interna no observable que es perturbada por algún ruido.
Por ejemplo, en el modelo lineal y gaussiano, las observaciones son de la forma:

$$y = \underbrace{\theta^\top x}_{\text{variable latente}} + \underbrace{\epsilon}_{\text{perturbación}}$$

En las clases anteriores se estudió como estimar los parámetros de un modelo. Ahora se verá cómo utilizar estas estimaciones para hacer predicciones. Para esto:

- Se considerará un modelo $p(y|x)$, con parámetro θ , del cual se han obtenido datos denotados mediante $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N$.
- Se considerará un modelo de **variable latente**: las observaciones consisten en una variable interna no observable que es perturbada por algún ruido.
Por ejemplo, en el modelo lineal y gaussiano, las observaciones son de la forma:

$$y = \underbrace{\theta^\top x}_{\text{variable latente}} + \underbrace{\epsilon}_{\text{perturbación}}$$

- De esta forma, al momento de hacer predicción, el objetivo será calcular la variable latente y no la observación futura ya que la observación tiene una perturbación.

Se denotará mediante \hat{f}_* e \hat{y}_* las predicciones de la variable latente f y la observación y para una nueva entrada x_* , condicional a los datos observados \mathcal{D} .

Se denotará mediante \hat{f}_* e \hat{y}_* las predicciones de la variable latente f y la observación y para una nueva entrada x_* , condicional a los datos observados \mathcal{D} .

Para un modelo lineal gaussiano con estimador puntual $\hat{\theta}$, el modelo corresponde a $y = \hat{\theta}^\top x + \epsilon$.

Se denotará mediante \hat{f}_* e \hat{y}_* las predicciones de la variable latente f y la observación y para una nueva entrada x_* , condicional a los datos observados \mathcal{D} .

Para un modelo lineal gaussiano con estimador puntual $\hat{\theta}$, el modelo corresponde a $y = \hat{\theta}^\top x + \epsilon$. Se tiene que:

- Predicción de la variable latente: $\hat{f}_* = \hat{\theta}^\top x_*$ (determinista).

Se denotará mediante \hat{f}_* e \hat{y}_* las predicciones de la variable latente f y la observación y para una nueva entrada x_* , condicional a los datos observados \mathcal{D} .

Para un modelo lineal gaussiano con estimador puntual $\hat{\theta}$, el modelo corresponde a $y = \hat{\theta}^\top x + \epsilon$. Se tiene que:

- Predicción de la variable latente: $\hat{f}_* = \hat{\theta}^\top x_*$ (determinista).
- Predicción de la observación: $y = \hat{\theta}^\top x_* + \epsilon$ (aleatoria).

Se denotará mediante \hat{f}_* e \hat{y}_* las predicciones de la variable latente f y la observación y para una nueva entrada x_* , condicional a los datos observados \mathcal{D} .

Para un modelo lineal gaussiano con estimador puntual $\hat{\theta}$, el modelo corresponde a $y = \hat{\theta}^\top x + \epsilon$. Se tiene que:

- Predicción de la variable latente: $\hat{f}_* = \hat{\theta}^\top x_*$ (determinista).
- Predicción de la observación: $y = \hat{\theta}^\top x_* + \epsilon$ (aleatoria).

Además, la aleatoriedad de ϵ permite representar esta predicción en términos de su esperanza y *barras de error*. Para el caso de ϵ gaussiano son explícitas, donde con un 95 % de probabilidad, $\hat{y}_* \in [\theta_{MV}^\top \tilde{x}_* - 2\sigma, \theta_{MV}^\top \tilde{x}_* + 2\sigma]$.

Por otro lado, cuando el parámetro θ es estimado de forma bayesiana, la posterior de θ es una distribución por lo que el modelo en sí es aleatorio.

Por otro lado, cuando el parámetro θ es estimado de forma bayesiana, la posterior de θ es una distribución por lo que el modelo en sí es aleatorio. De esta forma, se identifican dos fuentes de incertidumbre:

- Epistemológica: dada por la distribución posterior de θ .

Por otro lado, cuando el parámetro θ es estimado de forma bayesiana, la posterior de θ es una distribución por lo que el modelo en sí es aleatorio. De esta forma, se identifican dos fuentes de incertidumbre:

- Epistemológica: dada por la distribución posterior de θ .
- Aleatoria: dada por la distribución de ϵ .

Por otro lado, cuando el parámetro θ es estimado de forma bayesiana, la posterior de θ es una distribución por lo que el modelo en sí es aleatorio. De esta forma, se identifican dos fuentes de incertidumbre:

- Epistemológica: dada por la distribución posterior de θ .
- Aleatoria: dada por la distribución de ϵ .

Por lo tanto, para obtener una estimación de la variable latente, se debe integrar sobre las fuentes de incertidumbre (integrate-out):

$$\begin{aligned}\hat{f}_\star &\sim p(f_\star|x_\star, \mathcal{D}) \\ &= \int p(f_\star, \theta|x_\star, \mathcal{D})d\theta \\ &= \int p(f_\star|x_\star, \mathcal{D}, \theta)p(\theta|\mathcal{D}, x_\star)d\theta \\ &= \int p(f_\star|x_\star, \theta)p(\theta|\mathcal{D})d\theta\end{aligned}$$

Si bien el cálculo de la integral anterior puede ser complejo, se tienen dos casos particulares:

- **Caso gaussiano:** $p(f_*|x_*, \theta)p(\theta|\mathcal{D})$ corresponde al producto de dos gaussianas por lo que, de acuerdo al teorema de convolución, su integral es nuevamente una gaussiana.

Si bien el cálculo de la integral anterior puede ser complejo, se tienen dos casos particulares:

- **Caso gaussiano:** $p(f_\star|x_\star, \theta)p(\theta|\mathcal{D})$ corresponde al producto de dos gaussianas por lo que, de acuerdo al teorema de convolución, su integral es nuevamente una gaussiana.
- **Caso lineal:** basta notar que $f = \theta^\top \tilde{x}_\star$ y que $\theta \sim \mathcal{N}(\theta_N, \sigma^2 \Lambda_n^{-1})$. De esta forma, por linealidad:

$$\hat{f}_\star \sim p(f_\star|x_\star, \mathcal{D}) = \mathcal{N}(\theta_N^\top \tilde{x}_\star, \tilde{x}_\star^\top \sigma^2 \Lambda_n^{-1} \tilde{x}_\star)$$

Si bien el cálculo de la integral anterior puede ser complejo, se tienen dos casos particulares:

- **Caso gaussiano:** $p(f_\star|x_\star, \theta)p(\theta|\mathcal{D})$ corresponde al producto de dos gaussianas por lo que, de acuerdo al teorema de convolución, su integral es nuevamente una gaussiana.
- **Caso lineal:** basta notar que $f = \theta^\top \tilde{x}_\star$ y que $\theta \sim \mathcal{N}(\theta_N, \sigma^2 \Lambda_n^{-1})$. De esta forma, por linealidad:

$$\hat{f}_\star \sim p(f_\star|x_\star, \mathcal{D}) = \mathcal{N}(\theta_N^\top \tilde{x}_\star, \tilde{x}_\star^\top \sigma^2 \Lambda_n^{-1} \tilde{x}_\star)$$

Además, para determinar la predicción de la observación ruidosa $y = f + \epsilon$, solo se debe incorporar el estadístico del ruido:

$$\hat{y}_\star \sim \mathcal{N}(\theta_N^\top \tilde{x}_\star, \tilde{x}_\star^\top \sigma^2 \Lambda_n^{-1} \tilde{x}_\star + \sigma^2)$$

Si bien el cálculo de la integral anterior puede ser complejo, se tienen dos casos particulares:

- **Caso gaussiano:** $p(f_\star|x_\star, \theta)p(\theta|\mathcal{D})$ corresponde al producto de dos gaussianas por lo que, de acuerdo al teorema de convolución, su integral es nuevamente una gaussiana.
- **Caso lineal:** basta notar que $f = \theta^\top \tilde{x}_\star$ y que $\theta \sim \mathcal{N}(\theta_N, \sigma^2 \Lambda_n^{-1})$. De esta forma, por linealidad:

$$\hat{f}_\star \sim p(f_\star|x_\star, \mathcal{D}) = \mathcal{N}(\theta_N^\top \tilde{x}_\star, \tilde{x}_\star^\top \sigma^2 \Lambda_n^{-1} \tilde{x}_\star)$$

Además, para determinar la predicción de la observación ruidosa $y = f + \epsilon$, solo se debe incorporar el estadístico del ruido:

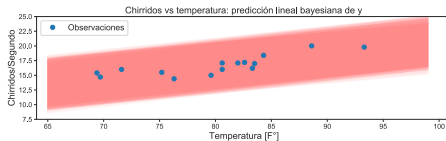
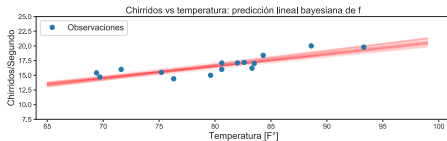
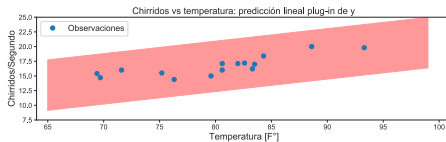
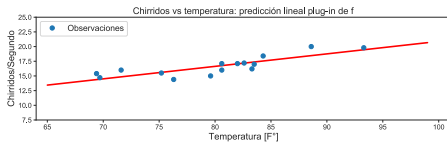
$$\hat{y}_\star \sim \mathcal{N}(\theta_N^\top \tilde{x}_\star, \tilde{x}_\star^\top \sigma^2 \Lambda_n^{-1} \tilde{x}_\star + \sigma^2)$$

Por último, a partir de la predicción bayesiana es posible obtener una predicción puntual dada por la media, donde para el caso lineal se tiene que:

$$\mathbb{E}[f_\star|x_\star, \mathcal{D}] = \mathbb{E}[\theta|\mathcal{D}]^\top \tilde{x}_\star = \bar{\theta}^\top \tilde{x}_\star$$

donde $\bar{\theta}$ corresponde a la media posterior de θ .

En la siguiente figura se ve la implementación de las 4 formas de realizar predicción:



El concepto de regresión lineal puede ser extendido mediante la aplicación de una transformación ϕ sobre la variable independiente x , construyendo un modelo lineal en la variable transformada $\phi = \phi(x)$ en lugar de en la variable original x .

El concepto de regresión lineal puede ser extendido mediante la aplicación de una transformación ϕ sobre la variable independiente x , construyendo un modelo lineal en la variable transformada $\phi = \phi(x)$ en lugar de en la variable original x .

Se considerarán transformaciones de la siguiente forma:

$$\begin{aligned}\phi: \mathbb{R}^M &\rightarrow \mathbb{R}^D \\ x &\mapsto \phi(x) = [\phi_1(x), \dots, \phi_D(x)]^\top\end{aligned}$$

donde $\phi_i: x \in \mathbb{R}^M \mapsto \phi_i(x) \in \mathbb{R}$ son funciones escalares $\forall i = 1, \dots, D$.

Regresión no lineal

El concepto de regresión lineal puede ser extendido mediante la aplicación de una transformación ϕ sobre la variable independiente x , construyendo un modelo lineal en la variable transformada $\phi = \phi(x)$ en lugar de en la variable original x .

Se considerarán transformaciones de la siguiente forma:

$$\begin{aligned}\phi: \mathbb{R}^M &\rightarrow \mathbb{R}^D \\ x &\mapsto \phi(x) = [\phi_1(x), \dots, \phi_D(x)]^\top\end{aligned}$$

donde $\phi_i: x \in \mathbb{R}^M \mapsto \phi_i(x) \in \mathbb{R}$ son funciones escalares $\forall i = 1, \dots, D$.

De esta forma, $\phi(x_i)$ puede representar las *características* de la observación cruda x_i .

Regresión no lineal

El concepto de regresión lineal puede ser extendido mediante la aplicación de una transformación ϕ sobre la variable independiente x , construyendo un modelo lineal en la variable transformada $\phi = \phi(x)$ en lugar de en la variable original x .

Se considerarán transformaciones de la siguiente forma:

$$\begin{aligned}\phi: \mathbb{R}^M &\rightarrow \mathbb{R}^D \\ x &\mapsto \phi(x) = [\phi_1(x), \dots, \phi_D(x)]^\top\end{aligned}$$

donde $\phi_i: x \in \mathbb{R}^M \mapsto \phi_i(x) \in \mathbb{R}$ son funciones escalares $\forall i = 1, \dots, D$.

De esta forma, $\phi(x_i)$ puede representar las *características* de la observación cruda x_i .

En la práctica, la función $\phi: x \mapsto \phi(x)$ es elegida en base al conocimiento *experto* que se tenga del problema. Nos referiremos a la construcción *manual* de la función ϕ como *ingeniería de características*.

Modelo lineal en los parámetros

Usando la nueva variable de características $\phi = \phi(x)$ se puede proponer un modelo lineal

$$y = \theta^\top \phi(x)$$

Modelo lineal en los parámetros

Usando la nueva variable de características $\phi = \phi(x)$ se puede proponer un modelo lineal

$$y = \theta^\top \phi(x)$$

Para un conjunto de observaciones $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N \subset \mathbb{R}^M \times \mathbb{R}$, este modelo puede ser entrenado con un funcional de costo cuadrático:

$$J = \sum_{i=1}^N (y_i - \theta^\top \phi(x_i))^2$$

Modelo lineal en los parámetros

Usando la nueva variable de características $\phi = \phi(x)$ se puede proponer un modelo lineal

$$y = \theta^\top \phi(x)$$

Para un conjunto de observaciones $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N \subset \mathbb{R}^M \times \mathbb{R}$, este modelo puede ser entrenado con un funcional de costo cuadrático:

$$J = \sum_{i=1}^N (y_i - \theta^\top \phi(x_i))^2$$

Además, se puede compactar el funcional utilizando la matriz de diseño:

$$\Phi = \begin{pmatrix} \phi(x_1)^\top \\ \vdots \\ \phi(x_N)^\top \end{pmatrix} = \begin{pmatrix} \phi_1(x_1) & \dots & \phi_D(x_1) \\ \vdots & \ddots & \vdots \\ \phi_1(x_N) & \dots & \phi_D(x_N) \end{pmatrix} \quad Y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}$$

Modelo lineal en los parámetros

Usando la nueva variable de características $\phi = \phi(x)$ se puede proponer un modelo lineal

$$y = \theta^\top \phi(x)$$

Para un conjunto de observaciones $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N \subset \mathbb{R}^M \times \mathbb{R}$, este modelo puede ser entrenado con un funcional de costo cuadrático:

$$J = \sum_{i=1}^N (y_i - \theta^\top \phi(x_i))^2$$

Además, se puede compactar el funcional utilizando la matriz de diseño:

$$\Phi = \begin{pmatrix} \phi(x_1)^\top \\ \vdots \\ \phi(x_N)^\top \end{pmatrix} = \begin{pmatrix} \phi_1(x_1) & \dots & \phi_D(x_1) \\ \vdots & \ddots & \vdots \\ \phi_1(x_N) & \dots & \phi_D(x_N) \end{pmatrix} \quad Y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}$$

Por lo que el funcional se reescribe como $J = \|Y - \Phi\theta\|_2^2$, cuyo mínimo es alcanzado en

$$\theta_\star = (\Phi^\top \Phi)^{-1} \Phi^\top Y.$$

Por último, es posible realizar una regularización sobre este modelo al igual que en MCR. Por ejemplo, para la regularización de ridge:

$$J_{\rho} = \|Y - \Phi\theta\|_2^2 + \rho \|\theta\|^2, \quad \rho \in \mathbb{R}^+.$$

Modelo lineal en los parámetros

Por último, es posible realizar una regularización sobre este modelo al igual que en MCR. Por ejemplo, para la regularización de ridge:

$$J_{\rho} = \|Y - \Phi\theta\|_2^2 + \rho \|\theta\|^2, \quad \rho \in \mathbb{R}^+.$$

En cuyo caso, es sabido que el funcional es minimizado en

$$\theta = (\Phi^{\top} \Phi + \rho \mathbb{I})^{-1} \Phi^{\top} Y$$

Modelo lineal en los parámetros

Por último, es posible realizar una regularización sobre este modelo al igual que en MCR. Por ejemplo, para la regularización de ridge:

$$J_\rho = \|Y - \Phi\theta\|_2^2 + \rho \|\theta\|^2, \quad \rho \in \mathbb{R}^+.$$

En cuyo caso, es sabido que el funcional es minimizado en

$$\theta = (\Phi^\top \Phi + \rho \mathbb{I})^{-1} \Phi^\top Y$$

Observación: la transformación afín-lineal $\tilde{x}_i = (x_i, 1)^\top$ usada al comienzo del curso, puede ser interpretada como un modelo de regresión no lineal bajo el mapa de características

$$x \in \mathbb{R}^M \mapsto \phi(x) = \begin{pmatrix} x \\ 1 \end{pmatrix} \in \mathbb{R}^{M+1}$$

Función Polinomial: $\phi = \{\phi_i\}_{i=0}^D$, donde $\phi_i(x) = x^i$, de tal forma que

$$\Phi = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^D \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_N & x_N^2 & \dots & x_N^D \end{bmatrix}.$$

Función Polinomial: $\phi = \{\phi_i\}_{i=0}^D$, donde $\phi_i(x) = x^i$, de tal forma que

$$\Phi = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^D \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_N & x_N^2 & \dots & x_N^D \end{bmatrix}.$$

- De acuerdo al teorema de Stone-Weierstrass, los polinomios son densos en $\mathcal{C}([a, b])$ (funciones continuas sobre compactos), por lo que es posible aproximar cualquier función continua mediante un polinomio.

Función Polinomial: $\phi = \{\phi_i\}_{i=0}^D$, donde $\phi_i(x) = x^i$, de tal forma que

$$\Phi = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^D \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_N & x_N^2 & \dots & x_N^D \end{bmatrix}.$$

- De acuerdo al teorema de Stone-Weierstrass, los polinomios son densos en $\mathcal{C}([a, b])$ (funciones continuas sobre compactos), por lo que es posible aproximar cualquier función continua mediante un polinomio.
- Una desventaja de esta base es que puede ser inestable: para obtener una buena aproximación polinomial, generalmente se requiere un grado D alto, por lo que los valores de $\phi(x)$ crecen, obviamente, de forma *polinomial*.

Función Polinomial: $\phi = \{\phi_i\}_{i=0}^D$, donde $\phi_i(x) = x^i$, de tal forma que

$$\Phi = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^D \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_N & x_N^2 & \dots & x_N^D \end{bmatrix}.$$

- De acuerdo al teorema de Stone-Weierstrass, los polinomios son densos en $\mathcal{C}([a, b])$ (funciones continuas sobre compactos), por lo que es posible aproximar cualquier función continua mediante un polinomio.
- Una desventaja de esta base es que puede ser inestable: para obtener una buena aproximación polinomial, generalmente se requiere un grado D alto, por lo que los valores de $\phi(x)$ crecen, obviamente, de forma *polinomial*.
- Por otra parte, la interpolación polinomial sufre del fenómeno de Runge, por lo que al utilizar un grado elevado, es posible que el error de predicción en los bordes crezca indefinidamente.

Función Sinusoidal: $\phi = \{\phi_i\}_{i=0}^D$, donde $\phi_i(x) = \cos\left(i\frac{2\pi}{2T}(x - b_i)\right)$, es decir:

$$\Phi = \begin{bmatrix} 1 & \cos\left(1\frac{2\pi}{2T}(x_1 - b_1)\right) & \dots & \cos\left(D\frac{2\pi}{2T}(x_1 - b_D)\right) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \cos\left(1\frac{2\pi}{2T}(x_N - b_1)\right) & \dots & \cos\left(D\frac{2\pi}{2T}(x_N - b_D)\right) \end{bmatrix}.$$

Función Sinusoidal: $\phi = \{\phi_i\}_{i=0}^D$, donde $\phi_i(x) = \cos\left(i\frac{2\pi}{2T}(x - b_i)\right)$, es decir:

$$\Phi = \begin{bmatrix} 1 & \cos\left(1\frac{2\pi}{2T}(x_1 - b_1)\right) & \dots & \cos\left(D\frac{2\pi}{2T}(x_1 - b_D)\right) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \cos\left(1\frac{2\pi}{2T}(x_N - b_1)\right) & \dots & \cos\left(D\frac{2\pi}{2T}(x_N - b_D)\right) \end{bmatrix}.$$

Una forma de evitar definir una fase, es considerar dos transformaciones por cada ϕ_i de la forma

$$\phi'_i(x) = \left(\sin\left(i\frac{2\pi}{2T}x\right), \cos\left(i\frac{2\pi}{2T}x\right) \right)$$

Función Sinusoidal: $\phi = \{\phi_i\}_{i=0}^D$, donde $\phi_i(x) = \cos\left(i\frac{2\pi}{2T}(x - b_i)\right)$, es decir:

$$\Phi = \begin{bmatrix} 1 & \cos\left(1\frac{2\pi}{2T}(x_1 - b_1)\right) & \dots & \cos\left(D\frac{2\pi}{2T}(x_1 - b_D)\right) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \cos\left(1\frac{2\pi}{2T}(x_N - b_1)\right) & \dots & \cos\left(D\frac{2\pi}{2T}(x_N - b_D)\right) \end{bmatrix}.$$

Una forma de evitar definir una fase, es considerar dos transformaciones por cada ϕ_i de la forma

$$\phi'_i(x) = \left(\sin\left(i\frac{2\pi}{2T}x\right), \cos\left(i\frac{2\pi}{2T}x\right) \right)$$

- Al igual que los polinomios, la base de senos y cosenos es también *universal* (más aún, forman una base de Hilbert de L^2 en el círculo).

Función Sinusoidal: $\phi = \{\phi_i\}_{i=0}^D$, donde $\phi_i(x) = \cos\left(i\frac{2\pi}{2T}(x - b_i)\right)$, es decir:

$$\Phi = \begin{bmatrix} 1 & \cos\left(1\frac{2\pi}{2T}(x_1 - b_1)\right) & \dots & \cos\left(D\frac{2\pi}{2T}(x_1 - b_D)\right) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \cos\left(1\frac{2\pi}{2T}(x_N - b_1)\right) & \dots & \cos\left(D\frac{2\pi}{2T}(x_N - b_D)\right) \end{bmatrix}.$$

Una forma de evitar definir una fase, es considerar dos transformaciones por cada ϕ_i de la forma

$$\phi'_i(x) = \left(\sin\left(i\frac{2\pi}{2T}x\right), \cos\left(i\frac{2\pi}{2T}x\right) \right)$$

- Al igual que los polinomios, la base de senos y cosenos es también *universal* (más aún, forman una base de Hilbert de L^2 en el círculo).
- Una desventaja de la base senoidal es que solo puede replicar funciones periódicas, con un período máximo en este caso de T .

Ejemplo de regresión no lineal

Consideremos el problema de predecir la cantidad de pasajeros en una aerolínea.

Ejemplo de regresión no lineal

Consideremos el problema de predecir la cantidad de pasajeros en una aerolínea. De forma incremental, se considerarán los siguientes mapas de características:

- polinomial.
- polinomial + senoidal (oscilatorio).
- polinomial + senoidal (oscilatorio) + amplitud creciente.

Ejemplo de regresión no lineal

Consideremos el problema de predecir la cantidad de pasajeros en una aerolínea. De forma incremental, se considerarán los siguientes mapas de características:

- polinomial.
- polinomial + senoidal (oscilatorio).
- polinomial + senoidal (oscilatorio) + amplitud creciente.

por lo tanto, denotando por x el tiempo e y la cantidad de pasajero, se considerará el siguiente modelo final:

$$y = \underbrace{\sum_{i=0}^3 \theta_i x^i}_{\text{parte polinomial}} + \underbrace{\sum_{i=1}^2 \alpha_i \exp(-\tau_i x^2) \cos(\omega_i(x - \psi_i))}_{\text{parte oscilatoria}}.$$

Ejemplo de regresión no lineal

Consideremos el problema de predecir la cantidad de pasajeros en una aerolínea. De forma incremental, se considerarán los siguientes mapas de características:

- polinomial.
- polinomial + senoidal (oscilatorio).
- polinomial + senoidal (oscilatorio) + amplitud creciente.

por lo tanto, denotando por x el tiempo e y la cantidad de pasajero, se considerará el siguiente modelo final:

$$y = \underbrace{\sum_{i=0}^3 \theta_i x^i}_{\text{parte polinomial}} + \underbrace{\sum_{i=1}^2 \alpha_i \exp(-\tau_i x^2) \cos(\omega_i(x - \psi_i))}_{\text{parte oscilatoria}}.$$

La motivación de este modelo es representar la tendencia de los datos mediante la componente polinomial y la oscilación anual mediante las componentes oscilatorias.

Ejemplo de regresión no lineal

Se cuenta con 12 años de datos con frecuencia mensual (144 datos), de los cuales solo 9 años (108 datos) han sido usado para encontrar los parámetros del modelo y los 3 años restantes (36 datos) para validar nuestras predicciones.

Ejemplo de regresión no lineal

Se cuenta con 12 años de datos con frecuencia mensual (144 datos), de los cuales solo 9 años (108 datos) han sido usado para encontrar los parámetros del modelo y los 3 años restantes (36 datos) para validar nuestras predicciones.

Los regresores obtenidos son los siguientes:

