Introducción GP

Rafael Molina
University of Granada (Spain)
http://decsai.ugr.es/~rms
rms@decsai.ugr.es

Índice

- La distribución de Gauss*
- De distribuciones gaussianas a procesos gaussianos
  - Funciones de covarianza
- Interpretación de un GP a través de un modelo generativo

* Recordatorio

I. La distribución de Gauss

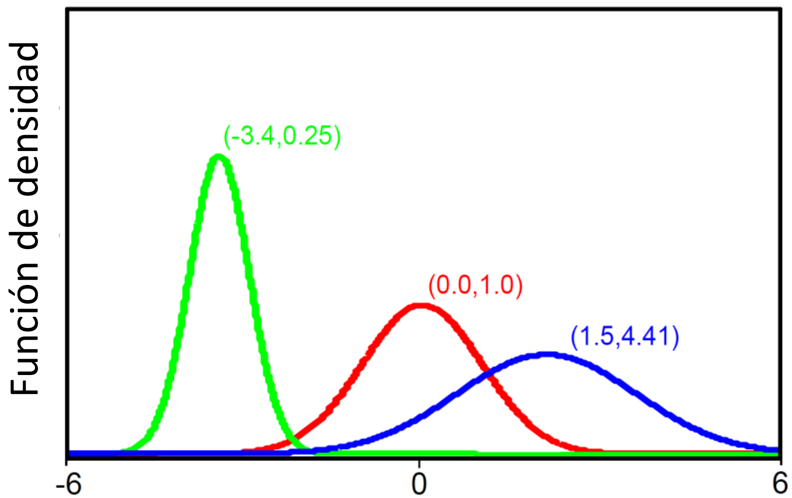
Caso Unidimensional

$$Pr(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-0.5(x-\mu)^2/\sigma^2]$$

Notación de Prince

que también escribiremos:

$$Pr(x|\mu,\sigma^2) = \mathrm{Norm}_x[\mu,\sigma^2] = \mathcal{N}(x|\mu,\sigma^2)$$



La distribución Normal Univariante describe una variable continua unidimensional.

Utiliza 2 parámetros μ y $\sigma^2>0$

I. La distribución de Gauss

Caso Multi-dimensional

D es el número de componentes de **x**

$$Pr(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma}|^{1/2}}\exp[-0.5(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})]$$

Notación de Prince

que también escribimos

$$Pr(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \mathrm{Norm}_{\mathbf{x}}[\boldsymbol{\mu},\boldsymbol{\Sigma}] = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma})$$

La distribución Normal Multivariante describe variables aleatorias múltiples. Utiliza dos parámetros:

- Un vector que contiene las medias, $\boldsymbol{\mu}$
- Una matriz de covarianzas simétrica “definida positiva” $\boldsymbol{\Sigma}$

Matriz definida positiva: $\mathbf{z}^T\boldsymbol{\Sigma}\mathbf{z}$ es positivo para cualquier vector no nulo $\mathbf{z}$

Transformación de Variables (muy importante)

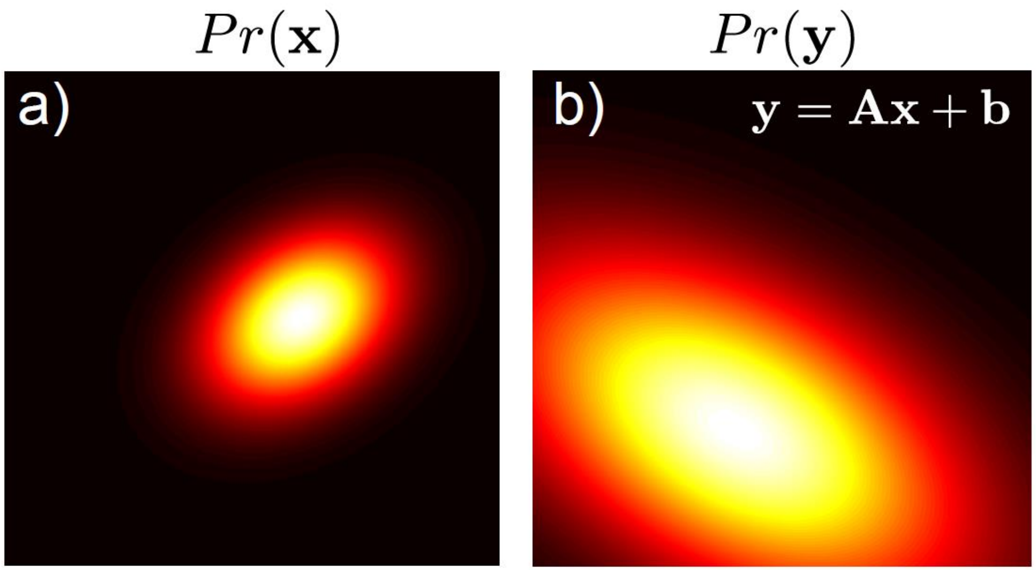
Si

$$Pr(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \text{Norm}_{\mathbf{x}}[\boldsymbol{\mu},\boldsymbol{\Sigma}]$$

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$$

entonces

$$Pr(\mathbf{y}) = \text{Norm}_{\mathbf{y}}[\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T)$$



Transformación de Variables (muy importante)

Si

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \boldsymbol{\epsilon}$$

donde

$$Pr(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathrm{Norm}_{\mathbf{x}}[\boldsymbol{\mu}, \boldsymbol{\Sigma}],$$
$$Pr(\boldsymbol{\epsilon}|\mathbf{0}, \sigma^2\mathbf{I}) = \mathrm{Norm}_{\boldsymbol{\epsilon}}[\mathbf{0}, \sigma^2\mathbf{I}]$$

siendo independientes $\mathbf{x}$ y $\boldsymbol{\epsilon}$

entonces

$$Pr(\mathbf{y}) = \mathrm{Norm}_{\mathbf{y}}[\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T + \sigma^2\mathbf{I}]$$

II. De Distribuciones Gaussianas a Procesos Gaussianos

Supongamos que las filas de **X** contienen los rasgos de nuestras muestras. Tenemos n muestras y cada una de ellas tiene q componentes

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,j} & \cdots & x_{1,q} \\ \vdots & \ddots & \vdots & \iddots & \vdots \\ x_{i,1} & \cdots & x_{i,j} & \cdots & x_{i,q} \\ \vdots & \iddots & \vdots & \ddots & \vdots \\ x_{n,1} & \cdots & x_{n,j} & \cdots & x_{n,q} \end{bmatrix} \rightarrow \mathbf{x}_{i,:}^{\top}$$

Fila completa

Rasgos de la i-ésima muestra

$$\mathbf{x}_{i,:} \in \mathcal{X} = \mathcal{R}^q$$

II. De Distribuciones Gaussianas a Procesos Gaussianos

Escribimos

$$p(\mathbf{f}|\mathbf{x}) = \mathcal{N}(\mathbf{f}|\mathbf{0}, \mathbf{K}_{ff}) = (2\pi)^{-\frac{n}{2}} |\mathbf{K}_{ff}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\mathbf{f}^\top \mathbf{K}_{ff}^{-1}\mathbf{f}\right)$$

Casi sería mejor utilizar aquí mayúscula para **x.**

La matriz de covarianzas

$$\mathbf{K}_{ff} = (k_f(\mathbf{x}_{i,:}, \mathbf{x}_{j,:}))$$

- se construye evaluando la función $\mathrm{k_f}$ sobre las parejas de muestras. Tiene tamaño nxn.
- Tiene que ser semidefinida positiva para cualquier número n y cualquier conjunto de vectores de rasgos.

II. De Distribuciones Gaussianas a Procesos Gaussianos

Observa que deberíamos escribir, incluyendo los parámetros desconocidos de la función de covarianza (núcleo),

$$p(\mathbf{f}|\mathbf{x}, \boldsymbol{\theta}_f)$$

Obviamente, una de las tareas durante el aprendizaje será aprender los parámetros del núcleo

II. De Distribuciones Gaussianas a Procesos Gaussianos

Ejemplo

$$\mathcal{X} = \mathcal{R}$$

$$\mathbf{X} = \begin{bmatrix} -2 \\ -1.8 \\ \vdots \\ 1.8 \\ 2 \end{bmatrix} \leftarrow \mathbf{x}_{2,:}$$

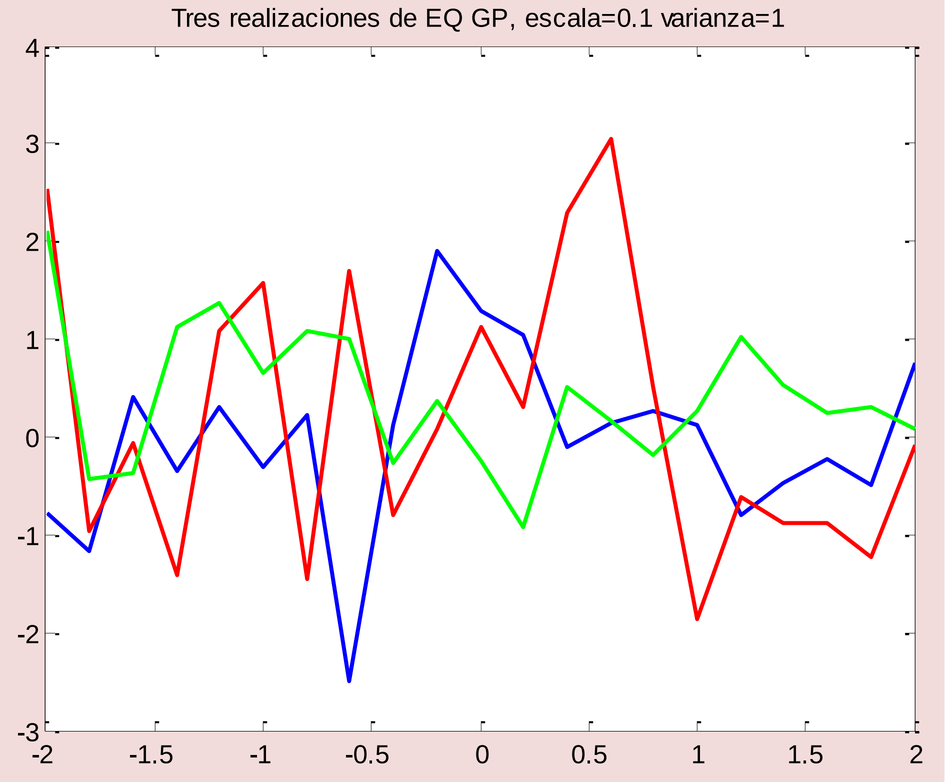
Fíjate que son números pero las filas podrían ser vectores.

$$\mathbf{K}_{ff} = (k_f(\mathbf{x}_{i,:}, \mathbf{x}_{j,:})) = (\mathrm{var} \times \exp[-\frac{1}{sc^2} \parallel \mathbf{x}_{i,:} - \mathbf{x}_{j,:} \parallel^2])$$

Exponentiated quadratic (EQ) covarianza. También se le llama RBF o núcleo Gaussiano

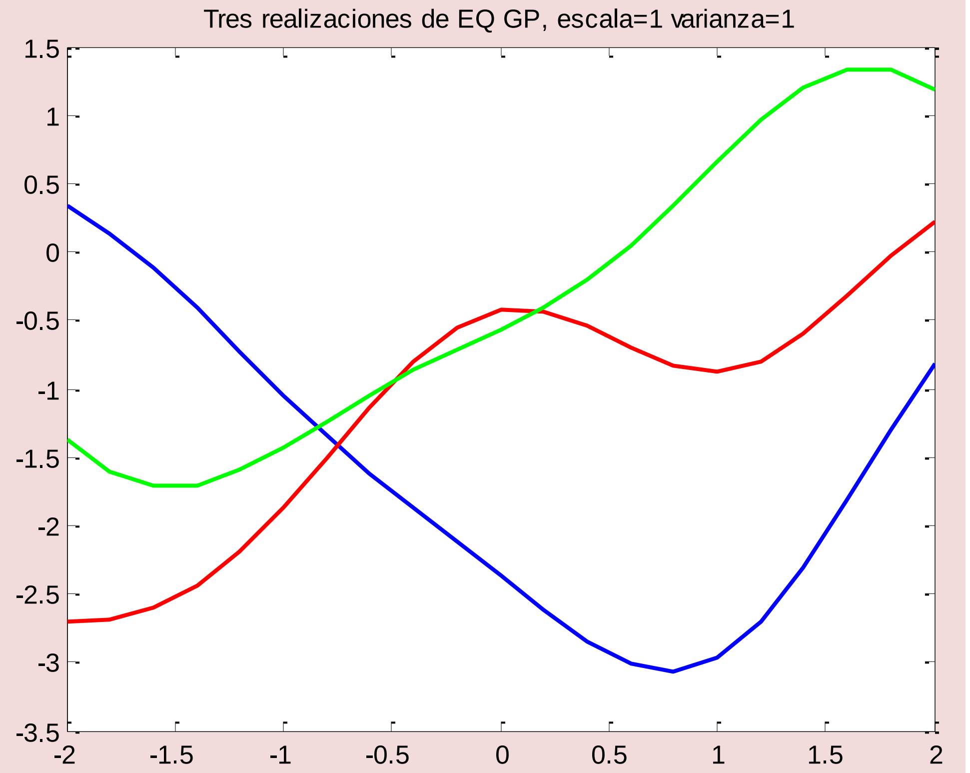
II. De Distribuciones Gaussianas a Procesos Gaussianos

```
% poco eficiente pero fácil de entender

sc=0.1; var=1;
X=-2:.2:2;
[U,V] = meshgrid(X,X);
Sigma = var*exp(-(1/(2*sc*sc))*(U-V).^2);
mean=zeros(1,numel(X));

sample=mvnrnd(mean,Sigma,1);
plot([-2:.2:2],sample,'b','LineWidth',2)

sample=mvnrnd(mean,Sigma,1);
hold on;plot([-2:.2:2],sample,'r','LineWidth',2)

sample=mvnrnd(mean,Sigma,1);
hold on; plot([-2:.2:2],sample,'g','LineWidth',2)

title(['Tres realizaciones de EQ GP, escala=',num2str(sc),' varianza=',num2str(var)])
```



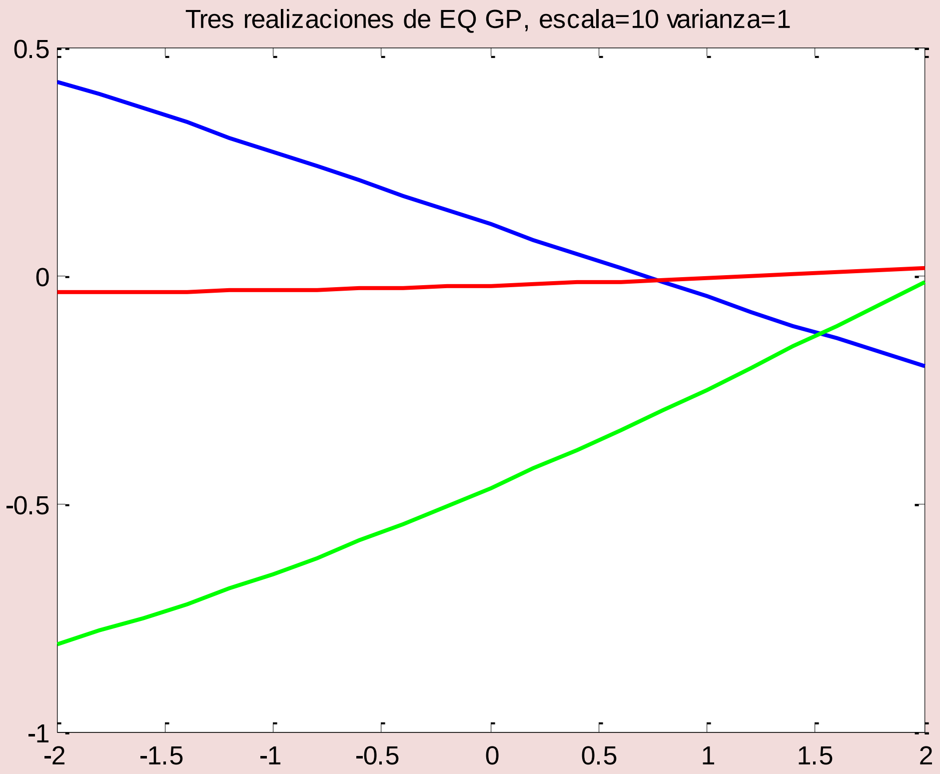
II. De Distribuciones Gaussianas a Procesos Gaussianos

```
% poco eficiente pero fácil de entender

sc=1; var=1;
X=-2:.2:2;
[U,V] = meshgrid(X,X);
Sigma = var*exp(-(1/(2*sc*sc))*(U-V).^2);
mean=zeros(1,numel(X));

sample=mvnrnd(mean,Sigma,1);
plot([-2:.2:2],sample,'b','LineWidth',2)

sample=mvnrnd(mean,Sigma,1);
hold on;plot([-2:.2:2],sample,'r','LineWidth',2)

sample=mvnrnd(mean,Sigma,1);
hold on; plot([-2:.2:2],sample,'g','LineWidth',2)

title(['Tres realizaciones de EQ GP, escala=',num2str(sc),' varianza=',num2str(var)]);
```



II. De Distribuciones Gaussianas a Procesos Gaussianos

```
% poco eficiente pero fácil de entender

sc=10; var=1;
X=-2:.2:2;
[U,V] = meshgrid(X,X);
Sigma = var*exp(-(1/(2*sc*sc))*(U-V).^2);
mean=zeros(1,numel(X));

sample=mvnrnd(mean,Sigma,1);
plot([-2:.2:2],sample,'b','LineWidth',2)

sample=mvnrnd(mean,Sigma,1);
hold on;plot([-2:.2:2],sample,'r','LineWidth',2)

sample=mvnrnd(mean,Sigma,1);
hold on; plot([-2:.2:2],sample,'g','LineWidth',2)

title(['Tres realizaciones de EQ GP, escala=',num2str(sc),' varianza=',num2str(var)]);
```



II. De Distribuciones Gaussianas a Procesos Gaussianos

En situaciones reales se supone que sólo disponemos de observaciones ruidosas y escribiremos

$$\mathbf{y} = f(\mathbf{x}) + \boldsymbol{\epsilon},\ \boldsymbol{\epsilon} \sim \mathcal{N}\left(\mathbf{0}, \beta^{-1}\mathbf{I}\right)$$

que induce la verosimilitud

$$p(\mathbf{y}|\mathbf{f}) = \mathcal{N}\left(\mathbf{y}|\mathbf{f}, \beta^{-1}\mathbf{I}\right)$$

y da lugar a la marginal (a utilizar para aprendizaje, training)

$$p(\mathbf{y}|\mathbf{x}) = \int p(\mathbf{y}|\mathbf{f})p(\mathbf{f}|\mathbf{x})\mathrm{d}\mathbf{f} = \mathcal{N}\left(\mathbf{y}|\mathbf{0}, \mathbf{K}_{ff} + \beta^{-1}\mathbf{I}\right)$$

que será utilizada para aprender los parámetros del modelo, es decir, los parámetros de la covarianza y β.

II. De Distribuciones Gaussianas a Procesos Gaussianos

La distribución predictiva (a utilizar para prueba, test) para una colección n_* de input de prueba x_*

$$p(\mathbf{f}_*|\mathbf{y}, \mathbf{x}, \mathbf{x}_*) = \mathcal{N}\left(\mathbf{f}_*|\mathbf{L}\mathbf{y}, \mathbf{K}_{**} - \mathbf{L}\mathbf{K}_{f*}\right)$$

donde

$$\mathbf{L} = \mathbf{K}_{*f}(\mathbf{K}_{ff} + \beta^{-1}\mathbf{I})^{-1}$$

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_{ff} & \mathbf{K}_{f*} \\ \mathbf{K}_{*f} & \mathbf{K}_{**} \end{bmatrix}$$

es la matriz de covarianza asociada a la matriz de rasgos de la derecha

$$\begin{bmatrix} \mathbf{X} \\ \mathbf{X}_* \end{bmatrix}$$

¿De dónde sale la fórmula?

II. De Distribuciones Gaussianas a Procesos Gaussianos

Demostración de la MUY IMPORTANTE fórmula de la distribución condicionada en la diapositiva anterior. Sabemos que

$$\mathrm{p}(\mathbf{f}_*|\mathbf{f}) = \mathcal{N}(\mathbf{f}_*|\mathbf{K}_{*f}\mathbf{K}_{ff}^{-1}\mathbf{f}, \mathbf{K}_{**} - \mathbf{K}_{*f}\mathbf{K}_{ff}\mathbf{K}_{f*})$$

$$\begin{aligned}\mathrm{p}(\mathbf{f}|\mathbf{y}) &= \mathcal{N}(\mathbf{f}|(\mathbf{K}_{ff}^{-1} + \beta\mathbf{I})^{-1}\beta\mathbf{y}, (\mathbf{K}_{ff}^{-1} + \beta\mathbf{I})^{-1}) \\ &= \mathcal{N}(\mathbf{f}|(\beta^{-1}\mathbf{K}_{ff}^{-1} + \mathbf{I})^{-1}\mathbf{y}, (\mathbf{K}_{ff}^{-1} + \beta\mathbf{I})^{-1})\end{aligned}$$

De donde obtenemos

$$\begin{aligned}\mathrm{p}(\mathbf{f}_*|\mathbf{y}) &= \int \mathrm{p}(\mathbf{f}_*|\mathbf{f})\mathrm{p}(\mathbf{f}|\mathbf{y})d\mathbf{f} \\ &= \mathcal{N}(\mathbf{f}_*|K_{*f}(K_{ff} + \beta^{-1}\mathbf{I})^{-1}\mathbf{y}, \\ &K_{**} - K_{*f}K_{ff}^{-1}K_{f*} + K_{*f}K_{ff}^{-1}(K_{ff}^{-1} + \beta^{-1}\mathbf{I})^{-1}K_{ff}^{-1}K_{f*}\end{aligned}$$

II. De Distribuciones Gaussianas a Procesos Gaussianos

Una de las grandes ventajas del uso de GP es la tratabilidad de su marginal

$$p(\mathbf{y}|\mathbf{x}) = \int p(\mathbf{y}|\mathbf{f})p(\mathbf{f}|\mathbf{x})\mathrm{d}\mathbf{f} = \mathcal{N}\left(\mathbf{y}|\mathbf{0}, \mathbf{K}_{ff} + \beta^{-1}\mathbf{I}\right)$$

$$p(\mathbf{y}|\mathbf{x}) = (2\pi)^{-\frac{n}{2}} \underbrace{\left|\mathbf{K}_{ff} + \beta^{-1}\mathbf{I}\right|^{-\frac{1}{2}}}_{\text{Nos gustaría que fuese grande}} \exp\left(-\frac{1}{2}\mathbf{y}^\top \underbrace{(\mathbf{K}_{ff} + \beta^{-1}\mathbf{I})^{-1}\mathbf{y}}_{\text{Nos gustaría que fuese pequeño}}\right)$$

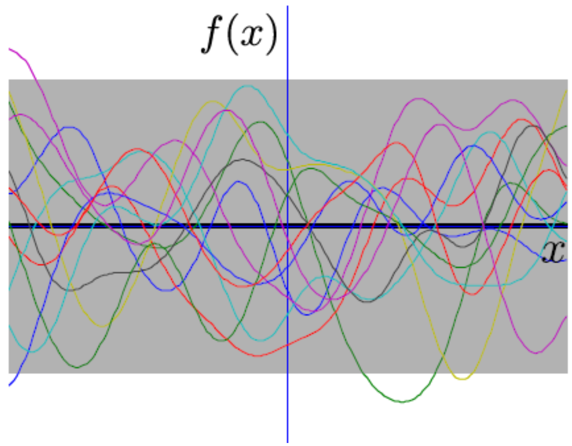
- La matriz de covarianza seleccionada decide las propiedades y no la forma paramétrica de f.
- El tamaño de la matriz de covarianzas nos generará problemas si tenemos que calcular su inversa.

II. De Distribuciones Gaussianas a Procesos Gaussianos

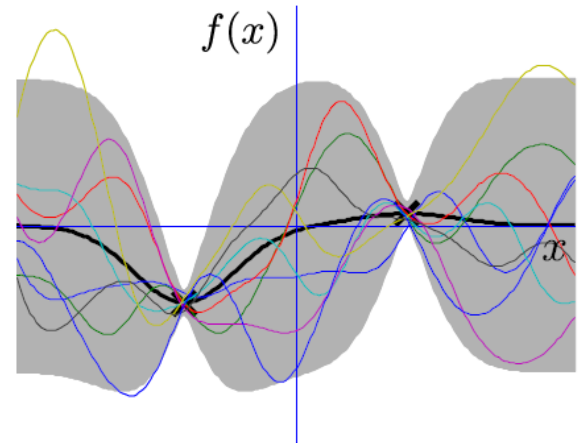
- La media del GP se representa como una curva en negro, dos desviaciones estándar se muestran con una sombra de gris.
- Las muestras se dibujan con diferentes colores.

Observa:

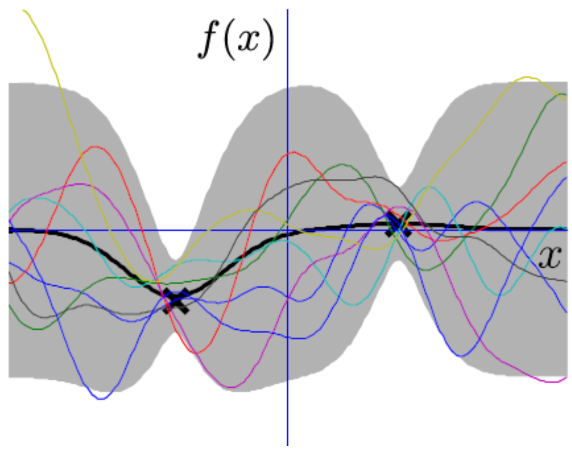
- Más ruido, más incertidumbre en la a posteriori, (b) y (c).
- Efecto de tener más observaciones, (d).
- La media vuelve a cero lejos de las observaciones.



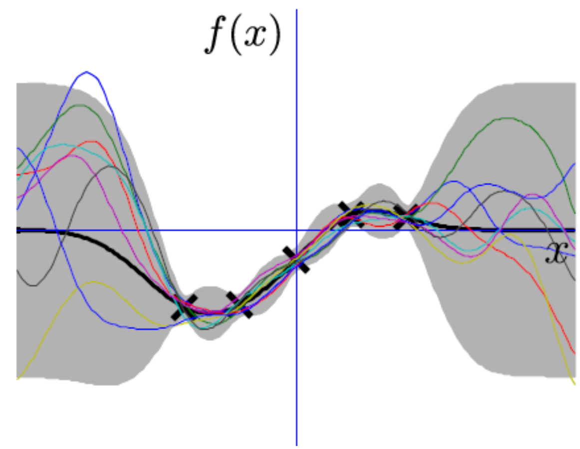
(a) Muestras de la a priori



(b) Muestras de la a posteriori



(c) Igual que (b) pero con más varianza del ruido



(d) Muestras de la a posteriori pero con más observaciones

II.1 Funciones de Covarianza

Vamos a ver varios ejemplos de matrices de covarianza.

Recuerda la notación

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,j} & \cdots & x_{1,q} \\ \vdots & \ddots & \vdots & \iddots & \vdots \\ x_{i,1} & \cdots & x_{i,j} & \cdots & x_{i,q} \\ \vdots & \iddots & \vdots & \ddots & \vdots \\ x_{n,1} & \cdots & x_{n,j} & \cdots & x_{n,q} \end{bmatrix} \rightarrow \mathbf{x}_{i,:}^\top$$

Fila completa

$\mathbf{x}_j$

Columna completa

$\mathbf{x}_{i,:}, \mathbf{x}_k,$

$k_f(\mathbf{x}_{i,:}, \mathbf{x}_{k,:})$ es semidefinida positiva sobre todos los posibles pares de entradas

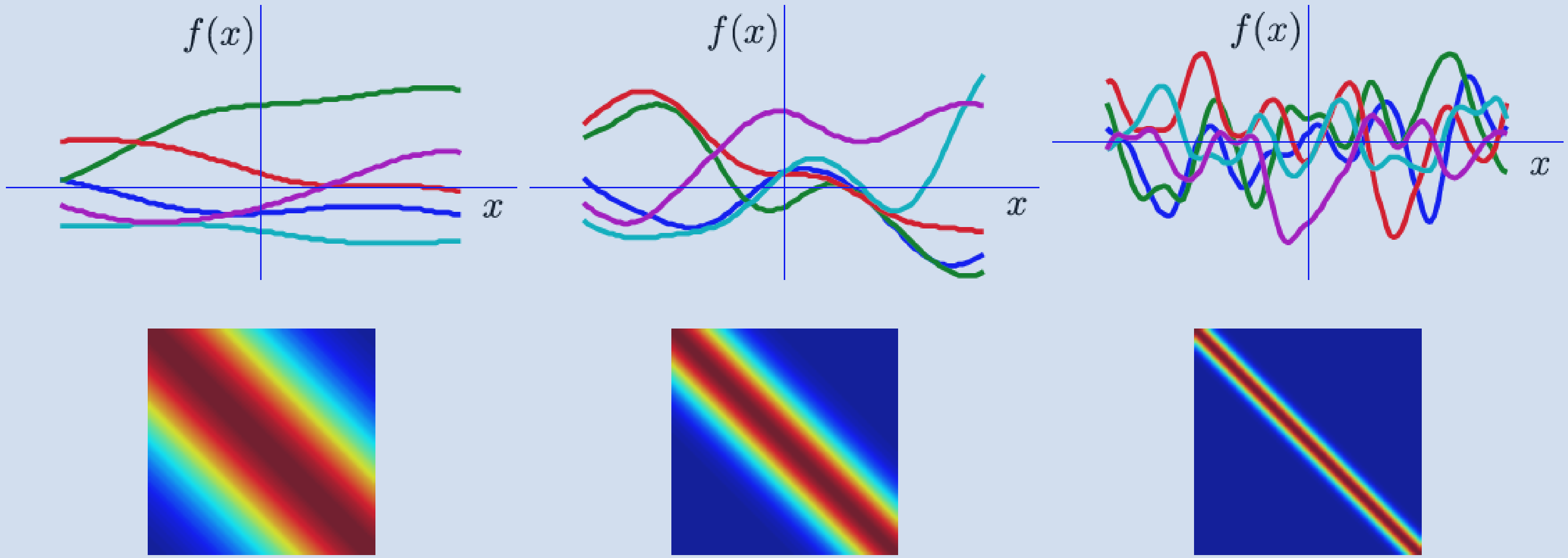
II.1 Funciones de Covarianza

Exponentiated quadratic (EQ), también llamada RBF.

$$k_{f(\mathrm{EQ})}\left(\mathbf{x}_{i,:}, \mathbf{x}_{k,:}\right) = \sigma_{\mathrm{EQ}}^{2} \exp\left(-\frac{1}{2\ell^{2}} \sum_{j=1}^{q} \left(x_{i,j} - x_{k,j}\right)^{2}\right)$$

Varianza y Longitud de escala (l^2)

Observa que $x_{i,:}$ pueden ser las coordenadas 2-D de un píxel



(a) Exp. Quadratic, $1/l^2 = 0.2$ (b) Exp. Quadratic, $1/l^2 = 1$ (c) Exp. Quadratic, $1/l^2 = 12$

Muestras con sus matrices de covarianza asociadas

II.1 Funciones de Covarianza

Exponentiated quadratic (EQ), con pesos diferentes para las distintas dimensiones de los inputs.

Permitirá Automatic Relevance Determination (ARD)

$$k_{f(\mathrm{ARD})}\left(\mathbf{x}_{i,:}, \mathbf{x}_{k,:}\right) = \sigma_{\mathrm{ARD}}^{2} \exp\left(-\frac{1}{2}\sum_{j=1}^{q} w_j \left(x_{i,j} - x_{k,j}\right)^2\right)$$

Pesos. Pequeños, dimensiones no relevantes.
Grandes, dimensiones relevantes

Lineal ARD función de covarianza.

$$k_{f(\mathrm{lin})}\left(\mathbf{x}_{i,:}, \mathbf{x}_{k,:}\right) = \sigma_{\mathrm{lin}}^{2}\, \mathbf{x}_{i,:}^{\top} \mathbf{C} \mathbf{x}_{k,:}$$

C es una matriz diagonal que contiene los pesos ARD.

II.1 Funciones de Covarianza

Matérn 3/2

$$k_{f(\text{mat})}\left(\mathbf{x}_{i,:},\mathbf{x}_{k,:}\right)=\sigma^2_{\text{mat}}\left(1+\frac{\sqrt{3}\left|\mathbf{x}_{i,:}-\mathbf{x}_{k,:}\right|}{\ell}\right)\exp\left(\frac{-\sqrt{3}\left|\mathbf{x}_{i,:}-\mathbf{x}_{k,:}\right|}{\ell}\right)$$

EQ-periodic

$$k_{f(\text{per})}\left(x_i,x_j\right)=\sigma^2_{\text{per}}\exp\left(-\frac{1}{\ell^2}2\sin^2\left(\frac{\pi}{T}\left|x_i-x_j\right|\right)\right)$$

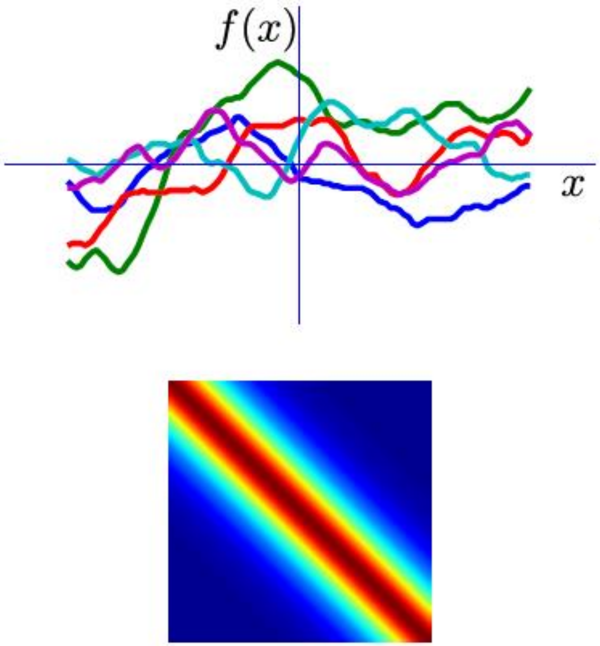
Se aplica a cada componente y luego el resultado se lleva a un EQ.

Ruido

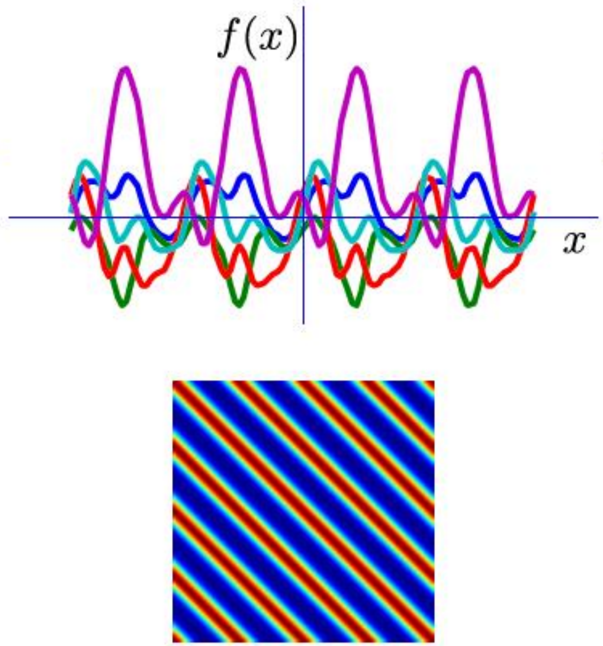
$$k_{\text{white}}\left(\mathbf{x}_{i,:},\mathbf{x}_{k,:}\right)=\theta_{\text{white}}\delta_{i,k}$$

Parece muy tonta pero aparece frecuentemente en problemas relacionados con diccionarios

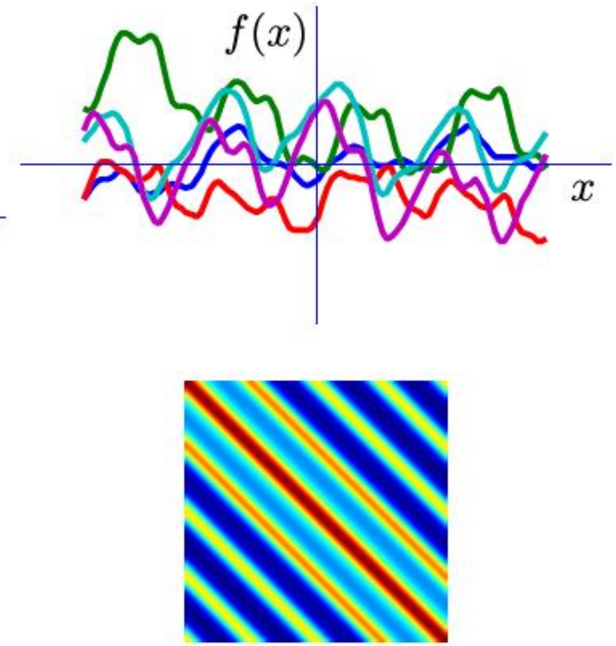
II.1 Funciones de Covarianza



(d) Matérn 3/2



(e) EQ-Periodic



(f) EQ plus EQ-Periodic

III Interpretación de un GP a través de un modelo generativo

Consideremos el modelo lineal

$$\mathbf{X} \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \beta^{-1}\mathbf{I}_{n\times n})$$

$$\mathbf{y} = \begin{bmatrix} f_1 \\ \vdots \\ f_n \end{bmatrix}_{n\times 1} = \boldsymbol{\mu}\mathbf{1}_{n\times 1} + \boxed{\begin{bmatrix} \mathbf{x}_{1,:}^{\mathrm{T}} \\ \vdots \\ \mathbf{x}_{n,:}^{\mathrm{T}} \end{bmatrix}_{n\times q}} \mathbf{w}_{q\times 1} + \boldsymbol{\epsilon}$$

y, muy importante, consideramos la distribución esférica

$$\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{q\times q})$$

III Interpretación de un GP a través de un modelo generativo

Si integramos en w y el error tenemos

$$\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}\mathbf{1}, \mathbf{X}\mathbf{X}^T + \beta^{-1}\mathbf{I})$$

III Interpretación de un GP a través de un modelo generativo

¿Y si hacemos la siguiente transformación?

$$\mathbf{x}_{i,:}^T = \left(\; x_{i1}, \;\; x_{i2}, \;\; \ldots, \;\; x_{iq} \; \right) \xrightarrow{\qquad \phi \qquad}$$

$$\mathbf{z}_{i,:}^T = \boldsymbol{\phi}(\mathbf{x}_{i,:}^T) = \left(\; \boldsymbol{\phi}_1(x_{i,:}^T) \;\; \boldsymbol{\phi}_2(x_{i,:}^T) \;\; \ldots \;\; \boldsymbol{\phi}_{q'}(x_{i,:}^T) \; \right)$$

Observa que las dimensiones de un vector de rasgos y su transformado pueden no ser las mismas. La dimensión q’ puede ser infinito!!!

III Interpretación de un GP a través de un modelo generativo

Podríamos hacer, por ejemplo, la siguiente transformación

$$\mathbf{x}^T = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \xrightarrow{\quad\quad\quad\quad\phi\quad\quad\quad\quad}$$

$$\mathbf{z} = \phi(\mathbf{x}^T) = \begin{pmatrix} x_1 & x_2 & x_1x_2 & x_1^2 & x_2^2 \end{pmatrix}$$

Observa que las dimensiones de un vector de rasgos y su transformado pueden no ser las mismas.

III Interpretación de un GP a través de un modelo generativo

Pasaríamos de

$$\mathbf{y} = \begin{bmatrix} f_1 \\ \vdots \\ f_n \end{bmatrix}_{n\times 1} = \boldsymbol{\mu}\mathbf{1}_{n\times 1} + \begin{bmatrix} \mathbf{x}_{1,:}^{\mathrm{T}} \\ \vdots \\ \mathbf{x}_{n,:}^{\mathrm{T}} \end{bmatrix}_{n\times q} \mathbf{w}_{q\times 1} + \boldsymbol{\epsilon}$$

$\mathbf{X}$

$\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{q\times q})$

$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{n\times n})$

a

$$\mathbf{y} = \begin{bmatrix} f_1 \\ \vdots \\ f_n \end{bmatrix}_{n\times 1} = \boldsymbol{\mu}'\mathbf{1}_{n\times 1} + \begin{bmatrix} \mathbf{z}_{1,:}^{\mathrm{T}} \\ \vdots \\ \mathbf{z}_{n,:}^{\mathrm{T}} \end{bmatrix}_{n\times q'} \mathbf{w}_{q'\times 1} + \boldsymbol{\epsilon}'$$

$\mathbf{Z}$

$\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{q'\times q'})$

$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, (\beta')^{-1}\mathbf{I}_{n\times n})$

III Interpretación de un GP a través de un modelo generativo

y si integramos en w y el nuevo error

$$\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}'\mathbf{1}, \mathbf{Z}\mathbf{Z}^T + (\beta')^{-1}\mathbf{I})$$

Este modelo nos indica que sólo necesitamos conocer el producto escalar de los vectores transformados en el dominio transformado.

La matriz $\mathbf{Z}\mathbf{Z}^{\mathrm{T}}$ es semidefinida positiva y corresponderá a una matriz de la forma

$$\mathbf{K}_{ff} = (k_f(\mathbf{x}_{i,:}, \mathbf{x}_{j,:}))$$

III Interpretación de un GP a través de un modelo generativo

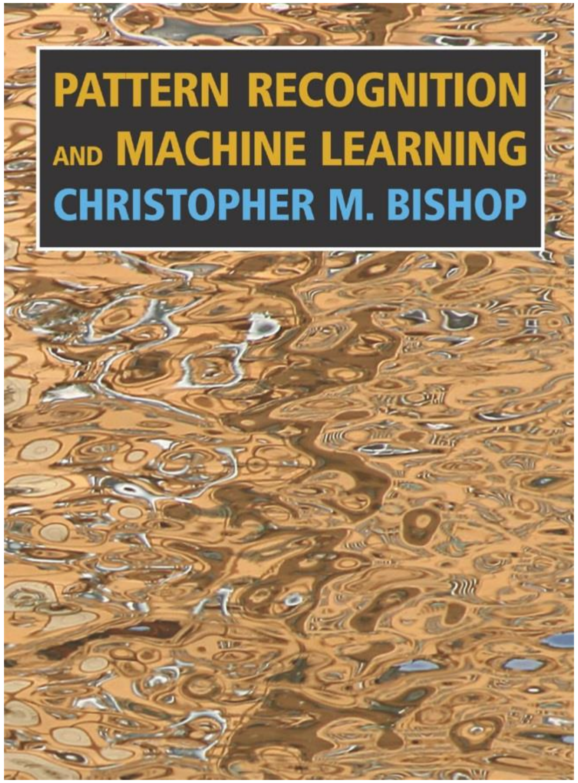
¿Qué hace, por tanto, un GP?

$$\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}'\mathbf{1}, \mathbf{Z}\mathbf{Z}^T + (\beta')^{-1}\mathbf{I})$$

- Como modelo generativo, representar linealmente nuestros datos en un dominio transformado de los rasgos.
- No necesitamos conocer en que vectores se transforman los rasgos. Sólo el producto escalar en el dominio transformado. Es decir, el núcleo asociado.

$$\mathbf{K}_{ff} = (k_f(\mathbf{x}_{i,:}, \mathbf{x}_{j,:}))$$

Bibliografía

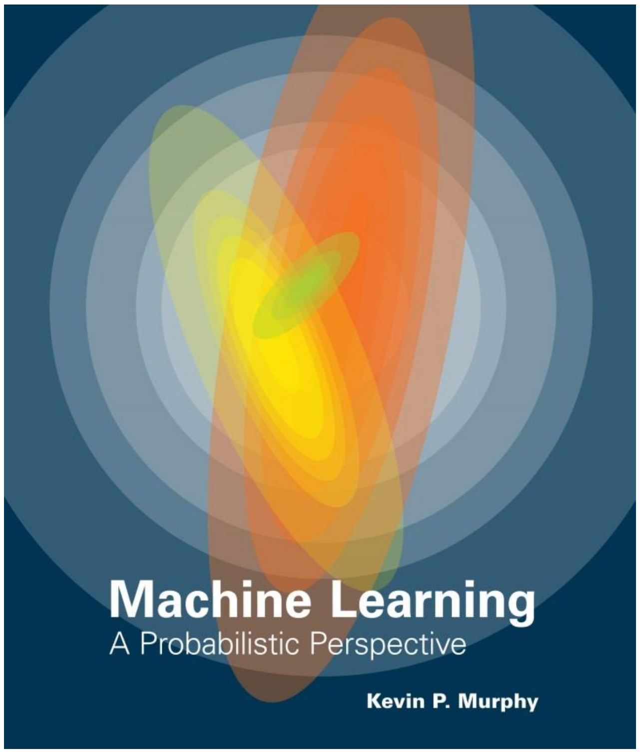


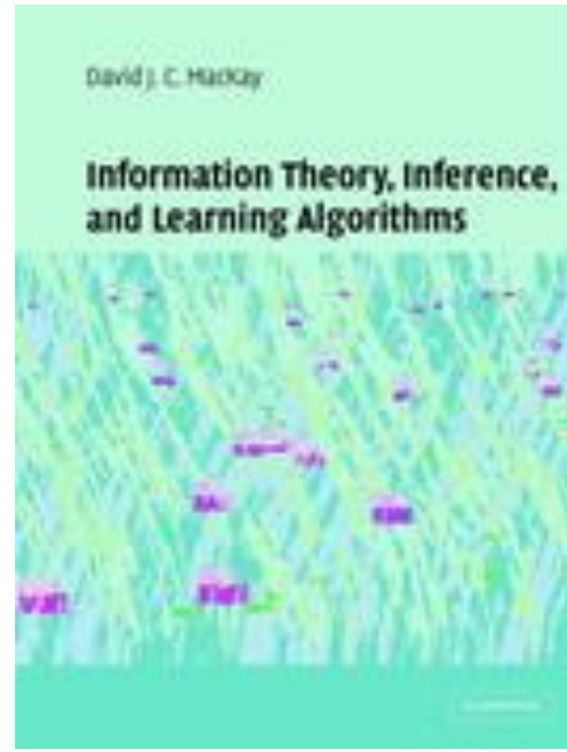
http://www.computervisionmodels.com/

http://research.microsoft.com/en-us/um/people/cmbishop/prml/

Bibliografía

http://www.cs.ubc.ca/~murphyk/MLbook/ http://www.inference.phy.cam.ac.uk/mackay/itila/





http://www.cs.ucl.ac.uk/staff/d.barber/brml/