Hilbert embeddings → parte 2.

Aplicaciones y estimadores.

Operador de covarianza cruzada:

$$C_{xy} = \mathbb{E}_{xy}\{\varphi(x)\otimes\phi(y)\} = \int \varphi(x)\otimes\phi(y)\, \partial P(x,y)$$

$$\tilde{C}_{xy} = \mathbb{E}_{xy}\{\varphi(x)\otimes\phi(y) - \mu_x\otimes\mu_y\}$$

Asumiendo $P(x,y) = P(x)P(y)$ y estimación del operador esperanza por media muestral $\left(\hat{P}(x,y) = \frac{1}{N}\right)$:

$$\hat{C}_{xy} = \frac{1}{N}\sum_{n=1}^{N}\varphi(x_n)\otimes\phi(y_n)\ ; \quad \text{desde i.i.d } \{x_n, y_n \in \mathcal{X}\}_{n=1}^{N}.$$

$$\bar{C}_{xy} = \frac{1}{N}\sum_{n=1}^{N}\varphi(x_n)\otimes\phi(y_n) - \hat{\mu}_x\otimes\hat{\mu}_y\ ; \quad \text{con}$$

$$\hat{\mu}_x = \frac{1}{N}\sum_{n=1}^{N}\varphi(x_n)\ ; \quad \hat{\mu}_y = \frac{1}{N}\sum_{n=1}^{N}\phi(y_n)$$

Con $H = I - \frac{1}{N}1_N^T 1_N$:

$$\bar{C}_{xy} = \frac{1}{N}\varphi_x H \phi_y\ ; \quad \begin{aligned}\varphi_x &= [\varphi(x_1)\cdots\varphi(x_N)]\\ \phi_y &= [\phi(y_1)\cdots\phi(y_N)]\end{aligned}$$

TAREA: Demostrar

Operador de covarianza cruzada como medida de dependencia.

Constraint Covarianza - COCO

$$\max_{f,g} \langle g, \hat{C}_{xy} f \rangle_{\mathcal{G}}$$

$$\text{s.t. } \|f\|_{\mathcal{F}} = 1 \ ; \ \|g\|_{\mathcal{G}} = 1.$$

$$f = \sum_{n=1}^{N} \alpha_n [\varphi(x_n) - \hat{\mu}_x] = \varphi_x H \alpha \ ; \quad \hat{\mu}_x = \frac{1}{N} \sum_{n=1}^{N} \varphi(x_n)$$

$$g = \sum_{m=1}^{N} \beta_m [\phi(y_n) - \hat{\mu}_y] = \phi_y H \beta \ ; \quad \hat{\mu}_y = \frac{1}{N} \sum_{m=1}^{N} \phi(y_n)$$

$$\mathcal{L}(f, g, \lambda, \gamma) = -f^T \hat{C}_{xy} g + \frac{\lambda}{2}(\|f\|_{\mathcal{F}}^2 - 1) + \frac{\gamma}{2}(\|g\|_{\mathcal{G}}^2 - 1)$$

$$f^T \hat{C}_{xy} g = \frac{1}{N} \alpha^T H \varphi_x^T (\varphi_x H \phi_y^T) \phi_y H \beta$$

$$= \frac{1}{N} \alpha^T \bar{K} \bar{L} \beta$$

NOTA: $H = HH$, además:

$$\|f\|_{\mathcal{F}}^2 = \alpha^T H \varphi_x \varphi_x^T H \alpha = \alpha^T \bar{K} \alpha$$

$$\mathcal{L} = -\frac{1}{N} \alpha^T \bar{K} \bar{L} \beta + \frac{\lambda}{2}(\alpha^T \bar{K} \alpha - 1) + \frac{\gamma}{2}(\beta^T \bar{L} \beta - 1)$$

NOTA: Similar a KCCA $\rightarrow \frac{\partial \mathcal{L}}{\partial \alpha} ; \frac{\partial \mathcal{L}}{\partial \beta}$

$$-\frac{1}{N}\bar{K}\bar{L}\beta + \lambda\bar{K}\alpha = 0 \;; \quad -\frac{1}{N}\bar{L}\bar{K}\alpha + \gamma\bar{L}\beta = 0$$

Como $\alpha^T\bar{K}\alpha = \beta^T\bar{L}\beta = 1$; $\lambda = \gamma$

$$\begin{bmatrix} 0 & \frac{1}{N}\bar{K}\bar{L} \\ \frac{1}{N}\bar{L}\bar{K} & 0 \end{bmatrix}\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \lambda \begin{bmatrix} \bar{K} & 0 \\ 0 & \bar{L} \end{bmatrix}\begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

valores propios generalizados $Av = \lambda Bv$.

Criterio de independencia Hilbert Schmidt

NOTA: Información Mutua $IM(X,Y) = \mathcal{E}_{xy}\left\{\log\left(\frac{p(x,y)}{p(x)p(y)}\right)\right\}$

$$IM(X,Y) = D_{KL}\left(p(x,y) \,\|\, p(x)p(y)\right)$$
$$= \mathcal{E}\{\log(p(x,y))\} - \mathcal{E}\{\log(p(x)p(y))\}$$

NOCIÓN DE INDEPENDENCIA. → Usando Hilbert-Emb.:

$$HSIC^2(\mathcal{F}, \mathcal{G}, P_{xy}) = \| C_{xy} - \mu_x \otimes \mu_y \|^2_{HS}$$

$$C_{xy} = \mathcal{E}\{\varphi(x) \otimes \phi(y)\} \;; \quad \mu_x = \mathcal{E}\{\varphi(x)\} \;; \quad \mu_y = \mathcal{E}\{\phi(y)\}$$

$$HSIC^2 = \|C_{xy}\|^2_{HS} - 2\langle C_{xy}, \mu_x \otimes \mu_y \rangle_{HS}$$
$$+ \|\mu_x \otimes \mu_y\|^2_{HS}$$

$$\|C_{xy}\|^2_{HS} = \langle C_{xy}, C_{xy}\rangle_{HS}$$
$$= \mathbb{E}_{xy}\mathbb{E}_{x'y'}\{\langle \varphi(x)\otimes\phi(y), \varphi(x')\otimes\phi(y')\rangle_{HS}\}$$
$$= \mathbb{E}_{xy}\mathbb{E}_{x'y'}\{\langle \varphi(x), \varphi(x')\rangle_{\mathcal{F}}\langle \phi(y), \phi(y')\rangle_{\mathcal{G}}\}$$
$$= \mathbb{E}_{xy}\mathbb{E}_{x'y'}\{k_x(x,x')\, k_y(y,y')\}$$

$$\langle \mu_x\otimes\mu_y, \mu_x\otimes\mu_y\rangle_{HS} = \langle \mu_x, \mu_x\rangle_{\mathcal{F}}\langle \mu_y, \mu_y\rangle_{\mathcal{G}}$$
$$= \mathbb{E}_x\mathbb{E}_{x'}\{k_x(x,x')\}\,\mathbb{E}_y\mathbb{E}_{y'}\{k_y(y,y')\}$$

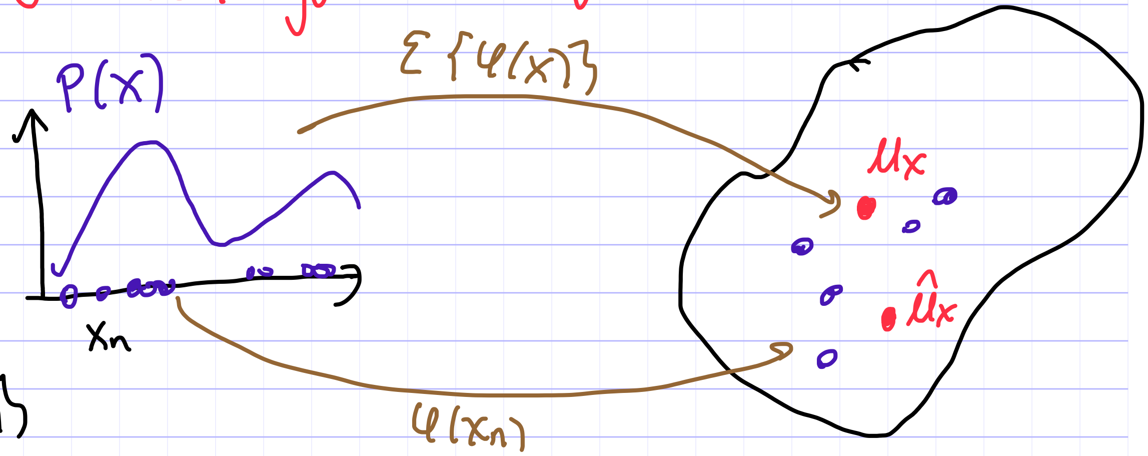
$$\langle C_{xy}, \mu_x\otimes\mu_y\rangle_{HS} = \mathbb{E}_{xy}\{\langle \varphi(x)\otimes\phi(y), \mu_x\otimes\mu_y\rangle_{HS}\}$$
$$= \mathbb{E}_{xy}\{\langle \varphi(x), \mu_x\rangle_{\mathcal{F}}\langle \phi(y), \mu_y\rangle_{\mathcal{G}}\}$$
$$= \mathbb{E}_{xy}\{\mathbb{E}_{x'}\{k_x(x,x')\}\,\mathbb{E}_{y'}\{k_y(y,y')\}\}$$

TAREA: compare HSIC con CKA y escriba HSIC en forma matricial.

Ej: $\|C_{xy}\|^2_{HS} = \frac{1}{N^2}\operatorname{tr}(KL)$

Hilbert Embeddings → condicionales:

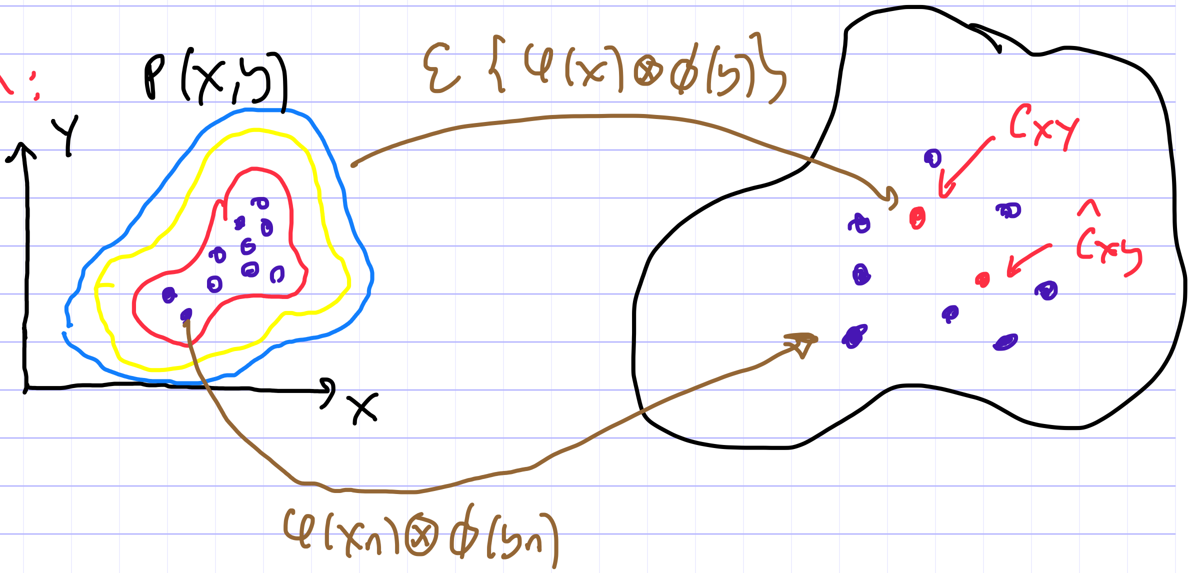
Marginal:



$$\mu_x = \mathcal{E}\{\varphi(x)\}$$

$$\hat{\mu}_x = \frac{1}{N}\sum_{n=1}^{N} \varphi(x_n)$$

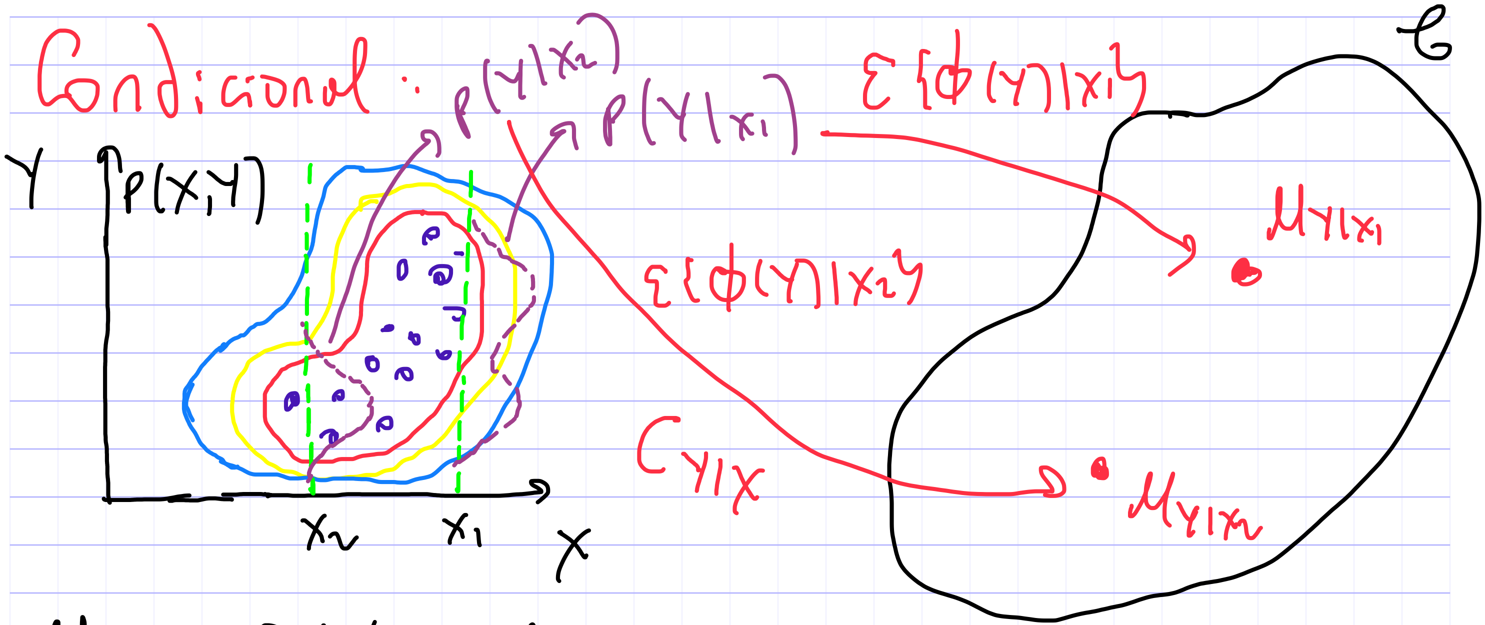
Conjunta:



$$C_{XY} = \mathcal{E}\{\varphi(x)\otimes\phi(y)\}$$

$$\hat{C}_{XY} = \frac{1}{N}\sum_{n=1}^{N} \varphi(x_n)\otimes\phi(y_n)$$

Condicional:



$$\mu_{Y|x} = \mathcal{E}\{\phi(Y)|x\}$$
$$= C_{Y|X}\varphi(x).$$

NOTA: Para el Hilbert Embedding condicional; se mapea una familia de puntos $\mu_{Y|x} \in \mathcal{G}$

$\mu_{Y|X} : \mathcal{F} \to \mathcal{G}$ OPERADOR. que satisface:

1. $\mu_{Y|x} = \mathcal{E}\{\phi(Y)|x\} = C_{Y|X}\varphi(x)$
2. $\mathcal{E}_{Y|x}\{g(Y)|x\} = \langle g, \mu_{Y|x}\rangle_{\mathcal{G}}$; $g \in \mathcal{G}$.

El operador condicional se puede definir a partir de operadores de covarianza cruzados como:

$$C_{XX}\mathcal{E}_{Y|X}\{g(Y)|X\} = C_{XY}\, g$$

se asume que $\mathcal{E}_{Y|X}\{g(Y)|X\} \in \mathcal{F}$.

Definición: $C_{Y|X} = C_{YX} C_{XX}^{-1}$. $(*)$

NOTA: Recuerda que estamos mapeando funciones de distribución a RKHSs:

$$P(Y,X) = P(Y|X)\,P(X) \rightarrow P(Y|X) = \frac{P(Y,X)}{P(X)}$$

Teorema Si $\mathcal{E}_{Y|X}\{g(Y)|X\} \in \mathcal{F}$, $C_{Y|X} = C_{YX} C_{XX}^{-1}$ cumple que: $\mu_{Y|x} = \mathcal{E}_{Y|x}\{\phi(Y)|x\}\varphi(x)$ y

$$\mathcal{E}_{Y|x}\{g(Y)|x\} = \langle g, \mu_{Y|x}\rangle_{G}$$

Prueba: Utilizando propiedad reproductiva:

$$\mathcal{E}_{Y|x}\{g(Y)|x\} = \langle \mathcal{E}_{Y|X}\{g(Y)|X\}, \varphi(x)\rangle_{\mathcal{F}}$$

dado que definimos $C_{XX}\mathcal{E}_{Y|X}\{g(Y)|X\} = C_{XY}\,g$

NOTA: Recuerda que en el espacio de entrada (sin mapeo a RKHSs):

$$C_{XY} = \mathcal{E}\{xy^T\}; \quad f^T C_{XY}\, g = \mathcal{E}_{XY}\{(f^T x)(g^T y)\}$$

$$g^T C_{YX} = g^T C_{Y|X} C_{XX} = \langle g^T C_{Y|X}, C_{XX}\rangle = \langle C_{XX}, g^T C_{Y|X}\rangle$$

$$g^T C_{YX} = C_{XX}\, g^T C_{Y|X} = C_{XY}\, g = C_{XX}\,\mathcal{E}_{Y|X}\{g^T Y|X\}$$

$$C_{XX}^{-1} C_{XY}\, g = \mathcal{E}_{Y|X}\{g^T Y|X\} = \mathcal{E}_{Y|X}\{g(Y)|X\} \in \mathcal{F}$$

Generalizando por propiedad reproductiva

Ahora para un $x \in \mathcal{X}$ y $\varphi : \mathcal{X} \to \mathcal{F}$ operador HS

$$\mathbb{E}_{Y|x}\{g(Y)|x\} = \langle \mathbb{E}_{Y|x}\{g(Y)|x\}, \varphi(x) \rangle_{\mathcal{F}} ;$$

dado que $\quad C_{XX}\mathbb{E}_{Y|X}\{g(Y)|X\} = C_{XY}\, g \quad$ y :

$$\mathbb{E}_{Y|x}\{g(Y)|X\} = C_{XX}^{-1} C_{XY}\, g = g^{T} C_{YX} C_{XX}^{-1}$$

$$\mathbb{E}_{Y|x}\{g(Y)|x\} \in \mathbb{R} = \langle C_{XX}^{-1} C_{XY}\, g , \varphi(x) \rangle_{\mathcal{F}}$$

Aplicando transpuesta

$$\mathbb{E}_{Y|x}\{g(Y)|x\} \in \mathbb{R} = \langle g, C_{YX} C_{XX}^{-1} \varphi(x) \rangle_{\mathcal{G}}$$

$$\mathbb{E}_{Y|x}\{g(Y)|x\} = \langle g, C_{Y|X}\, \varphi(x) \rangle_{\mathcal{G}} = \langle g, \mu_{Y|x} \rangle_{\mathcal{G}}$$

Estimador del Hilbert Embedding condicional desde conjunto de datos:

$\{x_n, y_n \in \mathcal{X}\}_{n=1}^{N}$; muestras i.i.d desde $P(X,Y)$,

y dado que:

$$C_{Y|X} = C_{YX} C_{XX}^{-1} ; \quad \mu_{Y|x} = C_{Y|X}\varphi(x) = C_{YX} C_{XX}^{-1} \varphi(x)$$

$$\hat{C}_{Y|X} = \Phi_Y (K_X + \lambda I)^{-1} \varphi_X^{T} \quad \text{TAREA: Demostrar}$$

$$\hat{\mu}_{Y|x} = \sum_{n=1}^{N} \beta_n(x) \phi(y_n) ; \qquad \beta(x) = [\beta_n(x)]_{n=1}^{N}$$

$$\beta(x) = (K_X + \lambda I)^{-1} k_X(x)$$

$$k_X(x) = [k(x, x_n)]_{n=1}^{N}$$

Sea $g(y) = \sum_{n=1}^{N} \alpha_n k(y_n, y)$

$$\mathbb{E}_{Y|x}\{g(Y)\} \simeq \langle g, \hat{\mu}_{Y|x} \rangle_{\mathcal{G}} = \sum_{n=1}^{N} \sum_{m=1}^{N} \alpha_n \beta_m(x) k(y_n, y_m)$$

$$\mathbb{E}_{Y|x}\{g(y)\} = \alpha^T K_y (K_x + \lambda I)^{-1} k_x.$$

NOTA: Evitamos aproximación de pdfs

Otras aplicaciones:

1. Maximum Mean Discrepancy = (MMD)

$$d(P(X), Q(Y)) = \| \mu_X - \mu_Y \|_F ; \quad x, y \in \mathcal{X}, \quad \varphi: \mathcal{X} \to F$$

$$\mu_X = \mathbb{E}_X\{\varphi(x)\} ; \quad \mu_Y = \mathbb{E}_Y\{\varphi(y)\}$$

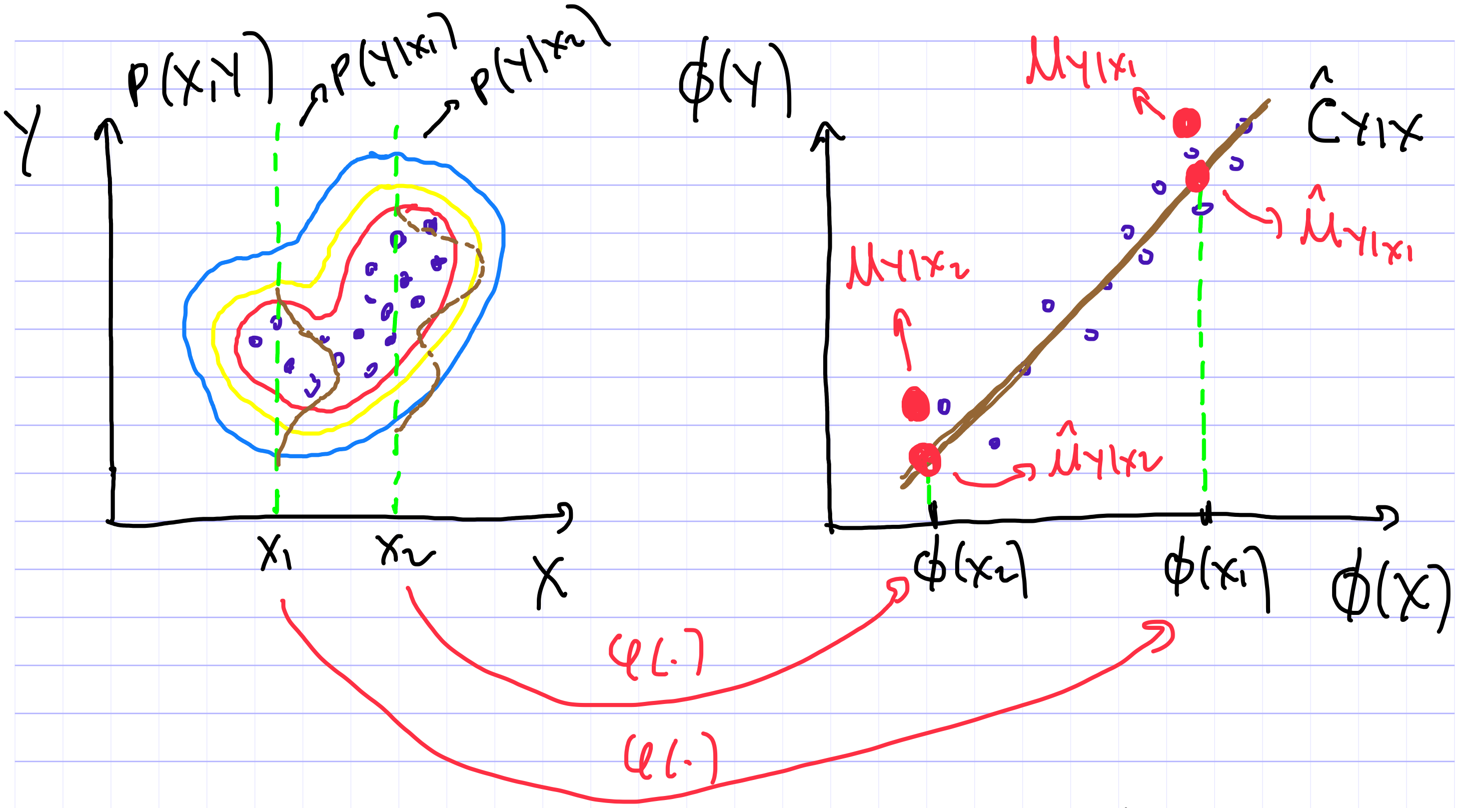
2. Mínimos cuadrados regularizados:

$$\hat{C}_{Y|X} = \underset{C: F \to F}{\arg\min}\ \mathbb{E}_{XY}\{ \| \phi(y_n) - C \phi(x_n) \|_F^2 \} + \ldots$$
$$\ldots + \lambda \| C \|_{HS}^2$$

$x, y \in \mathcal{X}$.

$\phi: \mathcal{X} \to F$.

TAREA: Encontrar estimadores matriciales de 1 y 2.
Implementar en tf2.0 para regresión.



$$\hat{\mu}_{Y|X} = \hat{C}_{Y|X}\,\phi(x) = \hat{C}_{YX}\left(\hat{C}_{XX}+\lambda I\right)^{-1}\phi(x)$$

$$\hat{\mu}_{Y|X} = \sum_{n=1}^{N} \beta_n(x)\,\phi(y_n)$$

Extensión a reglas probabilísticas

Regla de la suma

$$P(X) = \int P(X|Y)\,\partial P(Y)$$

$$\mu_X = \mathcal{E}_Y\{C_{X|Y}\,\phi(Y)\} = C_{X|Y}\,\mathcal{E}_Y\{\phi(Y)\} = C_{X|Y}\,\mu_Y$$

Regla de la cadena

$$P(X,Y) = P(X|Y)P(Y)$$

$$C_{XY} = \mathcal{E}_Y\{\mathcal{E}_{X|Y}\{\phi(X)\} \otimes \phi(Y)\}$$

$$C_{XY} = C_{X|Y}\,\mathcal{E}_Y\{\phi(Y) \otimes \phi(Y)\} = C_{X|Y}C_{YY}.$$

Regla de Bayes

$$P(Y|X) = P(X|Y)P(Y)/P(X)$$

$$P(X) = \int P(X|Y)\,\partial P(Y)$$

$$\mu_{Y|x} = C_{Y|X}\,\phi(x) = C_{YX}\,C_{XX}^{-1}\,\phi(x)$$