

Hilbert embeddings → parte 2.

Aplicaciones y estimadores.

Operador de covarianza cruzada:

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

$$\tilde{C}_{xy} = \mathbb{E}_{xy} \{ \varphi(x) \otimes \phi(y) - \bar{\mu}_x \otimes \bar{\mu}_y \}$$

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

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

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

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

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

$$\tilde{C}_{xy} = \frac{1}{N} \varphi_x H \phi_y; \quad \varphi_x = [\varphi(x_1) \dots \varphi(x_N)]$$

↗ $\phi_y = [\phi(y_1) \dots \phi(y_N)]$

TAREA: Demostrar

Operator de covarianza cruzada como medida de dependencia.

Constraint Covarianza - CCo

$$\max_{f, g} \langle g, \hat{C}_{xy} f \rangle_B$$

$$\text{s.t. } \|f\|_F = 1, \quad \|g\|_B = 1.$$

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

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

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

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

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

NOTA: $H = H^T$, además:

$$\|f\|_F^2 = \alpha^T H \Phi_x \Phi_x^T H \alpha = \alpha^T \bar{K} \alpha$$

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

NOTA: Similar a kCCA $\Rightarrow \frac{\partial \mathcal{L}}{\partial \alpha}; \frac{\partial \mathcal{L}}{\partial \beta}$

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

$$\text{Como } \alpha^T \bar{K} \alpha = \bar{B}^T \bar{B} = 1 ; \quad \lambda = \gamma$$

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

↑ Valores propios generalizadores $AU = \lambda BU$.

Criterio de independencia Hilbert Schmidt

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

$$\begin{aligned} IM(X,Y) &= D_{KL}(p(X,Y) || p(X)p(Y)) \\ &= \sum \{ \log(p(x,y)) \} - \sum \{ \log(p(x)p(y)) \} \end{aligned}$$

NOCIÓN DE INDEPENDENCIA. \rightarrow Usando Hilbert-Schmidt:

$$HSIC^2(f, g, p_{XY}) = \| C_{XY} - M_X \otimes M_Y \|_{HS}^2$$

$$C_{XY} = \{ \{ \varphi(x) \otimes \phi(y) \} ; \quad M_X = \{ \varphi(x) \} ; \quad M_Y = \{ \phi(y) \}$$

$$HSIC^2 = \| C_{XY} \|_{HS}^2 - 2 \langle C_{XY}, M_X \otimes M_Y \rangle_{HS}$$

$$+ \| M_X \otimes M_Y \|_{HS}^2$$

$$\begin{aligned}
\|C_{xy}\|_{HS}^2 &= \langle C_{xy}, C_{xy} \rangle_{HS} \\
&= \sum_{x,y} \sum_{x',y'} \left\{ \langle \varphi(x) \otimes \phi(y), \varphi(x') \otimes \phi(y') \rangle_{HS} \right\} \\
&= \sum_{x,y} \sum_{x',y'} \left\{ \langle \varphi(x), \varphi(x') \rangle_F \langle \phi(y), \phi(y') \rangle_E \right\} \\
&= \sum_{x,y} \sum_{x',y'} \left\{ K_x(x, x') K_y(y, y') \right\}
\end{aligned}$$

$$\begin{aligned}
\langle \mu_x \otimes \mu_y, \mu_x \otimes \mu_y \rangle_{HS} &= \langle \mu_x, \mu_x \rangle_F \langle \mu_y, \mu_y \rangle_E \\
&= \sum_x \sum_{x'} \left\{ K_x(x, x') \right\} \sum_y \sum_{y'} \left\{ K_y(y, y') \right\}
\end{aligned}$$

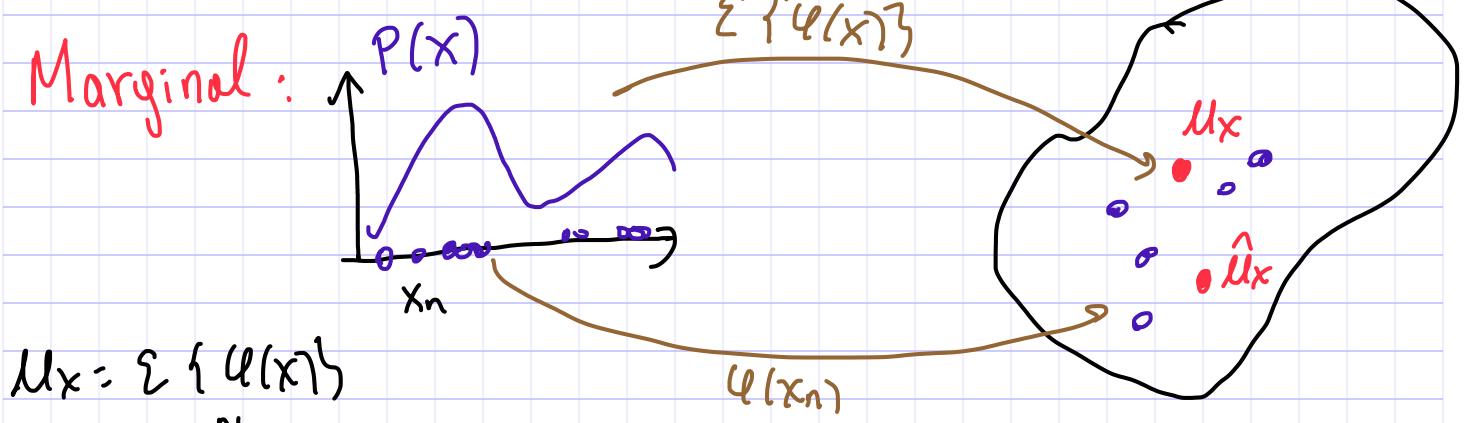
$$\begin{aligned}
\langle C_{xy}, \mu_x \otimes \mu_y \rangle_{HS} &= \sum_{x,y} \left\{ \langle \varphi(x) \otimes \phi(y), \mu_x \otimes \mu_y \rangle_{HS} \right\} \\
&= \sum_{x,y} \left\{ \langle \varphi(x), \mu_x \rangle_F \langle \phi(y), \mu_y \rangle_E \right\} \\
&= \sum_{x,y} \left\{ \sum_{x'} \left\{ K_x(x, x') \right\} \sum_{y'} \left\{ K_y(y, y') \right\} \right\}
\end{aligned}$$

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

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

Hilbert Embeddings → condicionales:

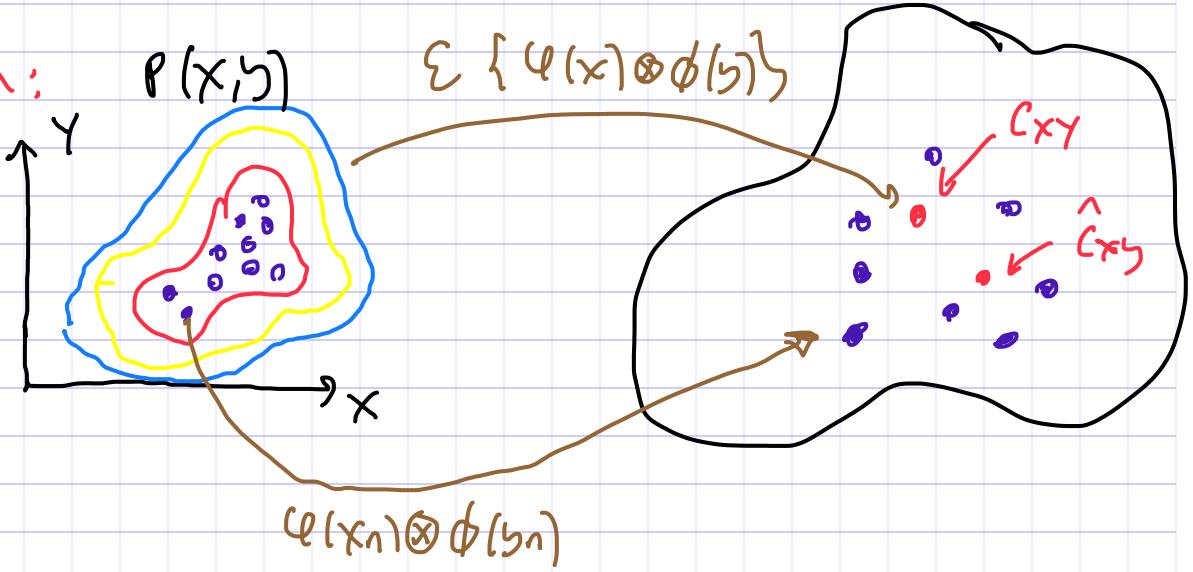
Marginal:



$$\mu_x = \sum \{ \varphi(x) \}$$

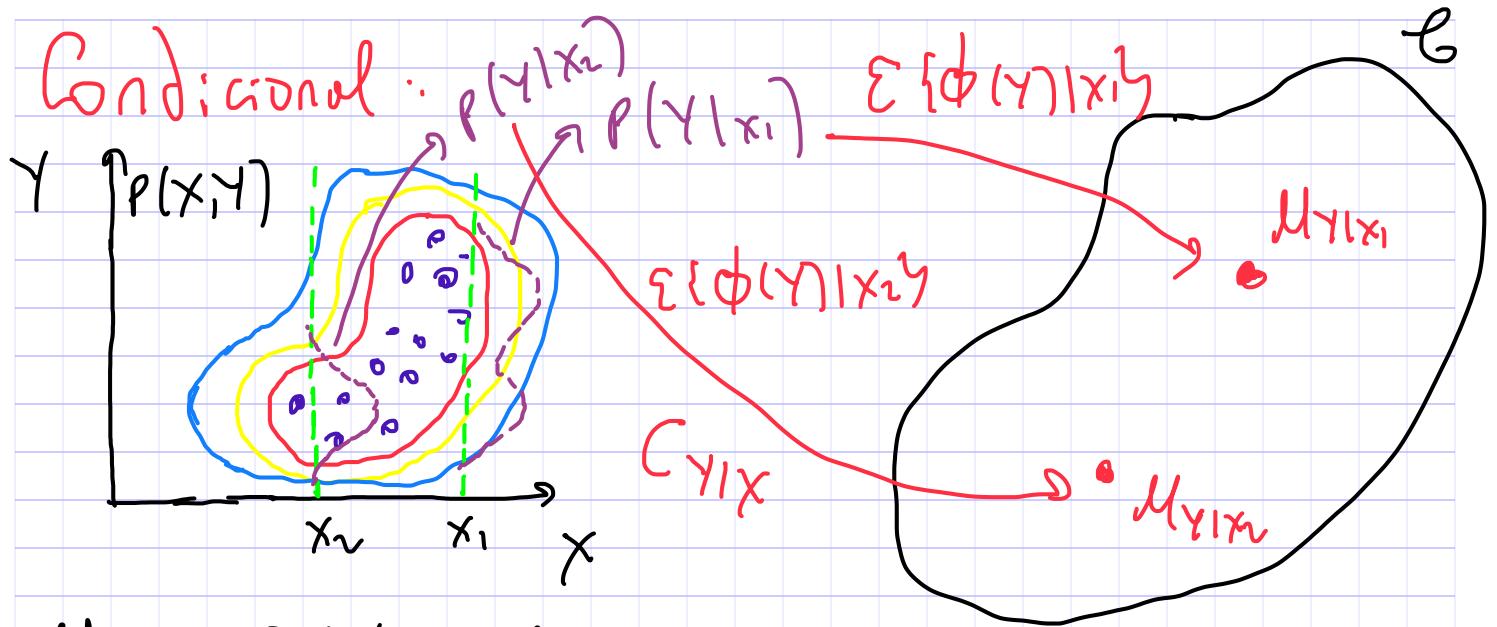
$$\hat{\mu}_x = \frac{1}{N} \sum_{n=1}^N \varphi(x_n)$$

Conjunta:



$$C_{xy} = \sum \{ \varphi(x) \otimes \phi(y) \}$$

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



$$\begin{aligned} \mu_{y|x} &= \{ \phi(y) | x \} \\ &= C_{y|x} \psi(x) . \end{aligned}$$

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

$$\mu_{y|x} : \mathcal{F} \rightarrow \mathcal{G} \quad \text{OPERADOR que satisface:}$$

$$1. \mu_{y|x} = \{ \phi(y) | x \} = C_{y|x} \psi(x)$$

$$2. \{ y|x \} \{ g(y) | x \} = \langle g, \mu_{y|x} \rangle_{\mathcal{G}} ; \quad g \in \mathcal{G}.$$

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

$$C_{xx} \{ y|x \} \{ g(y) | x \} = C_{xy} g$$

se asume que $\{ y|x \} \{ g(y) | x \} \in \mathcal{F}$.

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 $\mathbb{E}_{Y|X} \{ g(Y) | X \} \in F$, $C_{Y|X} = C_{YX} C_{XX}^{-1}$ cumple que: $M_{Y|X} = \mathbb{E}_{Y|X} \{ \phi(Y) | X \} \varphi(x)$ y
 $\mathbb{E}_{Y|X} \{ g(Y) | X \} = \langle g, M_{Y|X} \rangle_F$

Prueba: Utilizando propiedad reproductiva:

$$\mathbb{E}_{Y|X} \{ g(Y) | X \} = \langle \mathbb{E}_{Y|X} \{ g(Y) | X \}, \varphi(x) \rangle_F$$

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

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

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

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

$$g^T C_{YY} = C_{XX} \bar{g}^T C_{Y|X} = C_{XY} g = C_{XX} \mathbb{E}_{Y|X} \{ g^T Y | X \}$$

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

Generalizando por propiedad reproductiva)

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

$$\left\langle \gamma_{1x} \{ g(Y) | x \} \right\rangle = \left\langle \gamma_{1x} \{ g(Y) | x \}, \varphi(x) \right\rangle_{\mathbb{F}};$$

dado que $C_{xx} \gamma_{1x} \{ g(Y) | x \} = C_{xy} g$ y:

$$\gamma_{1x} \{ g(Y) | x \} = C_{xx}^{-1} C_{xy} g = g^T C_{yx} C_{xx}^{-1}$$

$$\left\langle \gamma_{1x} \{ g(Y) | x \} \right\rangle \in \mathbb{R} = \left\langle C_{xx}^{-1} (C_{xy} g), \varphi(x) \right\rangle_{\mathbb{F}}$$

Aplicando transpuesta

$$\left\langle \gamma_{1x} \{ g(Y) | x \} \right\rangle \in \mathbb{R} = \left\langle g, (C_{yx} C_{xx}^{-1} \varphi(x)) \right\rangle_{\mathbb{B}}$$

$$\left\langle \gamma_{1x} \{ g(Y) | x \} \right\rangle = \left\langle g, (\gamma_{1x} \varphi(x)) \right\rangle_{\mathbb{B}} = \left\langle g, \hat{\mu}_{Y|x} \right\rangle_{\mathbb{B}}$$

Estimador del Hilbert Embedding condicional
dado conjunto de datos:

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

$$\gamma_{1x} = C_{yx} C_{xx}^{-1}; \quad \hat{\mu}_{Y|x} = \gamma_{1x} \varphi(x) = C_{yx} C_{xx}^{-1} \varphi(x)$$

$$\hat{\gamma}_{1x} = \Phi_y (K_x + \lambda I)^{-1} \varphi_x^T \quad \xrightarrow{\text{TAREA: Demostrar}}$$

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

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

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

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

$$E_{\gamma|x} \{g(\gamma)\} \approx \langle g, \hat{\mu}_{\gamma|x} \rangle_B = \sum_{n=1}^N \sum_{m=1}^N \alpha_n \beta_m(x) k(y_n, \gamma_m)$$

$$E_{\gamma|x} \{g(\gamma)\} = \alpha^T K_y (K_x + \lambda I)^{-1} \gamma.$$

NOTA: Evitamos aproximación de pdfs \uparrow

Otras aplicaciones:

1. Minimum Mean Discrepancy = (MMD)

$$\delta(P(x), Q(y)) = \| \mu_x - \mu_y \|_F; \quad x, y \in \mathcal{X}$$

$\varphi: \mathcal{X} \rightarrow \mathcal{F}$

$$\mu_x = \mathbb{E}_x \{ \varphi(x) \}; \quad \mu_y = \mathbb{E}_y \{ \varphi(y) \}$$

2. Mínimos cuadrados regularizados:

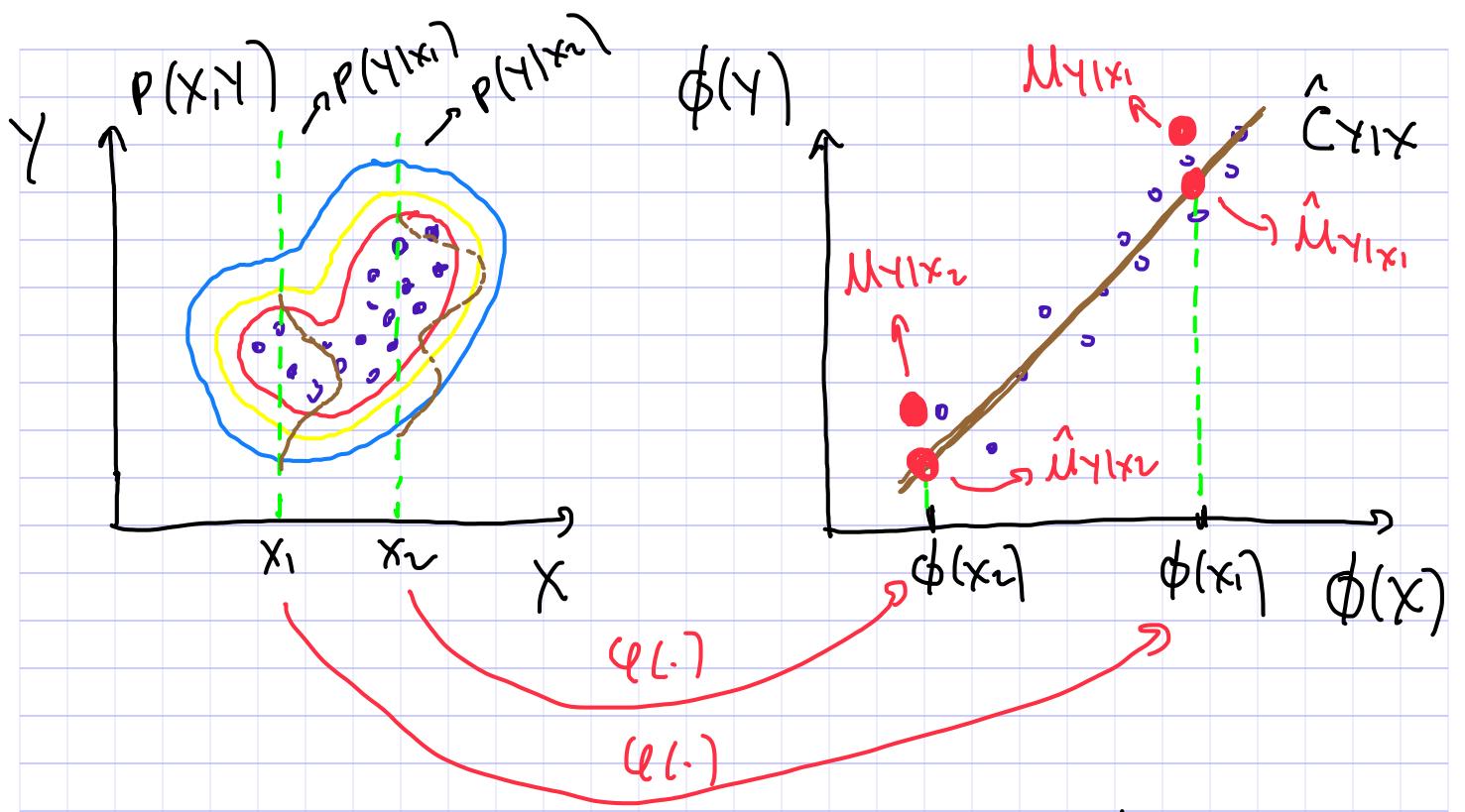
$$\hat{\gamma}_{|x} = \arg \min_{C: F \rightarrow F} E_{xy} \left\{ \| \phi(y_n) - C \phi(x_n) \|_F^2 \right\}$$

$$x \| C \|_{HS}^2$$

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

$$\phi: \mathcal{X} \rightarrow \mathcal{F}$$

TAREA: Encontrar estimadores matriciales de γ_1 y γ_2 .
Implementar en tensorflow para regresión.



$$\hat{M}_{y|x} = \hat{C}_{yx} \phi(x) = \hat{C}_{yx} (\hat{C}_{xx} + \lambda I)^{-1} \phi(x)$$

$$\hat{M}_{y|x} = \sum_{n=1}^N B_n(x) \phi(y_n)$$