Motivacion supongamos un conjunto de variables de entrada o covariantes, z, y una salida x. La idea seria ver como depende la variable de salida x con los covariates. En general se quisiera estimar la densidad condicional $p(x|z)$.

A partir de los datos que son un conjunto de n pares, (x_i, z_i), los objetivos son:

- Poder evaluar analiticamente (parametricamente?) a $p(x|z)$.
- Generar un conjunto de muestras x_j a partir de $p(x|z^*)$ dado un valor del target z^*.

Buscamos un mapa que elimine de la densidad de x toda la variabilidad explicable por z. Esto es lo que hacemos cuando obtenemos $p(x|z)$ dado un z^*.

Para encontrar ese mapa pongo dos condiciones

$$\min_{y=T(x,z)} \mathbb{E}_\pi[c(x,y)] \tag{1}$$

sujeto a que y sea independiente de z. En algun sentido diria a que no quede mas informacion de x en z.

Una vez que encontre el T lo puedo usar para generar muestras de la condicional:

Dado un z^* y dado los datos $\{x_i, z_i\}_{1:n}$ puedo generar las $y_i = T(x_i, z_i)$ y luego antitransformo a $x_i^* = T^{-1}(y_i, z^*)$ esto me daria datos posibles de la variabilidad de x dado z^*.

uedo hacer esto con el MPF-Stein? Transformo

1 Formulación Dual de Kantorovich

- Problema Primal
$$\min_\pi \int \int c(x,y)\, d\pi(x,y) \tag{2}$$
sujeto a que π tenga marginales $\rho_1, \rho_2, \ldots, \rho_K$.
- Problema Dual
$$\max_{\{\phi_k\}_{k=1}^K} \sum_{k=1}^K \int \phi_k(x)\rho_k(x)\, dx \tag{3}$$

Para el caso general de K medidas:

$$\sum_{k=1}^K \phi_k(x_k) \leq c(x_1, x_2, \ldots, x_K) \quad \forall (x_1, \ldots, x_K) \tag{4}$$

Interpretación

- $\phi_k(x)$: potenciales de Kantorovich
- $\int \phi_k(x)\rho_k(x)\, dx$: valor total de la masa en ρ_k evaluada con precios ϕ_k
- La maximización busca los precios que maximizan el beneficio total respetando las restricciones de arbitraje

Dualidad Fuerte (Teorema de Kantorovich)

Bajo condiciones regulares:

$$\min_{\pi} \int \int c(x,y)\, d\pi(x,y) = \max_{\phi_1,\phi_2} \left\{ \int \phi_1(x)\rho_1(x)\, dx + \int \phi_2(y)\rho_2(y)\, dy \right\} \tag{5}$$

2 Kantorovich

$$\min_{\mu,\pi_k} \sum_{k=1}^{K} P_k \int c(x,y)\pi_k(x,y)\, dx\, dy \tag{6}$$

$$\int \pi_k(x,y)\, dy = \rho_k(x), \quad \int \pi_k(x,y)\, dx = \mu(y) \tag{7}$$

Where: - μ is the barycenter distribution (optimization variable) - $\pi_k(x,y)$ are the joint distributions coupling ρ_k and μ (optimization variables) - P_k are the weights for each source distribution - $c(x,y)$ is the cost function (squared Euclidean distance in this case) - $\rho_k(x)$ are the source distributions - The first constraint ensures π_k has ρ_k as its x-marginal - The second constraint ensures π_k has μ as its y-marginal

Formulación de Kantorovich - Todas las Ecuaciones

Problema Primal Relajado

Función Objetivo:

$$\min_{\mu,\pi_k} \sum_{k=1}^{K} P_k \int c(x,y)\pi_k(x,y)\, dx\, dy \tag{8}$$

Restricciones:

$$\int \pi_k(x,y)\, dy = \rho_k(x), \quad \int \pi_k(x,y)\, dx = \mu(y) \tag{9}$$

Problema Dual

Función Objetivo Dual:

$$\max_{\phi_k,\psi_k} \sum_{k=1}^{K} \int \phi_k(x)\rho_k(x)\, dx \tag{10}$$

Restricciones Duales:

$$\forall x,y: \quad \phi_k(x) + \psi_k(y) \leq P_k c(x,y), \quad \forall y: \sum_{k=1}^{K} \psi_k(y) \geq 0 \tag{11}$$

Relación entre Dual y Primal

$$Y_k(x) = x - \frac{1}{P_k}\nabla\phi_k(x) \tag{12}$$

Simplificación del Dual

Restricciones Equivalentes:

$$\forall \mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_K \end{pmatrix}, \quad \sum_{k=1}^{K} \phi_k(x_k) \leq \min_{y} \sum_{k=1}^{K} P_k c(x_k, y) \tag{13}$$

Función de Costo Independiente de y:

$$C(x_1, \ldots, x_K) = \min_y \sum_{k=1}^{K} P_k c(x_k, y) \tag{14}$$

Problema Dual Simplificado:

$$\max_{\phi_k} \sum_{k=1}^{K} \int \phi_k(x)\rho_k(x)\, dx \tag{15}$$

sujeto a:

$$\forall \mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_K \end{pmatrix}, \quad \sum_{k=1}^{K} \phi_k(x_k) \leq C(x_1, \ldots, x_K) \tag{16}$$

Relación Primal-Dual para Distancia Cuadrática

Para la función de costo de distancia cuadrática, la solución óptima tiene la forma:

$$\pi_k(x, y) = \rho_k(x)\delta(y - Y_k(x)) \tag{17}$$

3 Implementacion

$$\min_{y_i = T(x_i, z_i)} \max_{g,f} \frac{1}{n} \sum_{i=1}^{n} \left[c(x_i, y_i) + \lambda\, g(y_i) f(z_i) \right], \quad \text{subject to} \quad \sum_{i=1}^{n} f(z_i) = 0, \quad \|f\| = \|g\| = 1 \tag{18}$$

Primeramente necesitamos encontrar las g y f que maximizan la funcion de costo. Para esto vamos a restringir la busqueda de las funciones f y g a espacios de dimension finita ($\mathcal{F}$ y $\mathcal{G}$). En lugar de considerar todas las funciones posibles, solo consideramos combinaciones lineales de un conjunto finito de funciones base (”features” o características).

$$f(z) = F(z)a = \sum_{j=1}^{m_z} F^j(z)\, a_j, \qquad g(y) = G(y)b = \sum_{k=1}^{m_y} G^k(y)\, b_k \tag{19}$$

donde: $F(z)$ y $G(y)$ son vectores de funciones base predefinidas (por ejemplo, polinomios, funciones trigonométricas, kernels, etc.). $a \in \mathbb{R}^{m_z}$ y $b \in \mathbb{R}^{m_y}$ son los vectores de coeficientes a determinar. m_z y m_y son las dimensiones de estos espacios de características, elegidas por el usuario.

Se reemplaza $G(y)$ por un operador ortogonal $Q_y(y)$ que genere el mismo rango efectivo a traves de SVD,

$$\mathbf{G}_0 = \mathbf{U\Sigma V}^T \tag{20}$$

donde G_0 es la matriz evaluada en los puntos y_{0j}.

La cantidad de dimensiones de $Q_y(y)$ es tal que genere gran parte de la variabilidad con m_y tal que $\sum_k^{m_y} \sigma_k^2 > 0.99$

Se define una nueva matriz de transformación $\mathbf{B}^y$ de tamaño $m_y \times n_y$ que proyecta los coeficientes originales al espacio reducido y ortogonal:

$$\mathbf{B}^y_{jk} = \frac{1}{\sigma_k}\mathbf{V}_{jk}, \quad \text{para } j = 1, \ldots, m_y;\ k = 1, \ldots, n_y \tag{21}$$

El nuevo operador de características ortogonales Q_y se define como:

$$Q_y(y) = G(y)\,\mathbf{B}^y \tag{22}$$

Entonces hemos representado al problema de maximizacion de funciones en un espacio finito y a traves de coeficientes que tienen que ser determinados a traves de la maximizacion.

Entonces ahora el problema min-max lo tenemos en a, b, y_l,

$$\min_{\{y_i\}} \left\{ \max_{a,b} \sum_{i=1}^{n} c(x_i, y_i) + \lambda \sum_{h=1}^{n_z} \sum_{l=1}^{n_y} \left(\sum_{i=1}^{n} Q^h_z(z_i) Q^l_y(y_i) \right) a_h b_l, \quad \|a\| = \|b\| = 1 \right\} \tag{23}$$

Se puede determinar en forma explicita la maximizacion en esencia a y b se deben alinear con las primeras componentes principales a izquierda y derecha.

Nuevamente entonces hacemos SVD sobre la matriz $\mathbf{A} = \mathbf{Q}_z^\top \mathbf{Q}_y = \mathbf{U\Sigma V}^\top$.

Entonces terminamos en un problema de minimizacion dado por

$$\min_{\{y_i\}} \sum_{i=1}^{n} c(x_i, y_i) + \lambda \, \|A(y)\| \tag{24}$$

podemos hacer un problema de descenso de gradientes dado por:

$$y_i^{n+1} = y_i^n - \eta^n \left[\frac{1}{n} \nabla_y c(x_i, y)\big|_{y_i^n} + \lambda\, a' \, \nabla_y A\big|_{y_i^n} b \right] \tag{25}$$

donde $\nabla_y A^{hl}\Big|_{y_i^n} = Q^h_z(z_i) \sum_j \nabla G^j(y)\big|_{y_i} B^y_{jl}$