Motivacion

Supongamos un conjunto de variables de entrada, z, y una salida x.
La idea seria ver como depende la variable de salida x con las variables de entrada z.
En general este problema consiste en inferir 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 $p(x|z)$.
- Dado un valor del target z^*, generar un conjunto de muestras x_j a partir de $p(x|z^*)$

Problema de transporte de Monge del baricentro

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

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

sujeto a que y sea independiente de z.

En algun sentido decimos a que no quede mas informacion de x en z.

Densidad condicional: Metodo generativo con OTBP

Una vez que encontre el T lo puedo usar para generar muestras de la condicional.
Dado los datos $\{x_i, z_i\}_{1:n}$ puedo transformar las muestras $y_i = T(x_i, z_i)$.
Si tengo un z^* sobre el cual quiero generar muestras de x condicionadas a z^*
Antitransformo sobre x a $x_i^* = T^{-1}(y_i, z^*)$ esto me dar N muestras posibles de la variabilidad de x dado z^*.

Problema primal relajado

Función Objetivo

$$\min_{\mu,\pi} \int c(x,y)\pi(x,y)\,dx\,dy$$

Restricciones

$$\int \pi(x,y)\,dy = \rho(x), \quad \int \pi(x,y)\,dx = \mu(y)$$

donde $\mu(y)$ es la distribucion del baricentro y la π es la distribucion conjunta que acopla ρ con μ

Problema de adversarios empirico

Dadas las muestras $(x_i, z_i)_{i=1}^N$, proponemos trabajar en un espacios de fucniones $\mathcal{F}$ para z y $\mathcal{G}$ para y

$$\min_{y_i = T(x_i, z_i)} \max_{g,f} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left[c(x_i, y_i) + \lambda\, g(y_i) f(z_i) \right] \right\},$$

Restricciones: $\sum_{i=1}^n f(z_i) = 0, \quad \|f\| = \|g\| = 1$

Primeramente necesitamos encontrar las g y f en el espacio de funciones que maximizan la funcion de costo.

La independencia impone una condicion de “sin correlación” entre la g y la f. Si y y z son independientes se tiene que

$$\mathbb{E}_\pi [g(y) f(z)] = 0$$

Test functions

Vamos a restringir la busqueda de las funciones f y g a espacios de dimension finita ($\mathcal{F}$ y $\mathcal{G}$).
Solo consideramos combinaciones lineales de un conjunto finito de funciones base ("features" o características).

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

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.

Transformo al espacio ortogoanal

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

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

donde $\mathbf{G}_0$ 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:

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

Problema de adversarios en el espacio ortogonal

El nuevo operador de características ortogonales Q_y se define como:

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

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,

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

Determinacion explicita del a y b que maximizan

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
$\mathbf{a}^\top \mathbf{A} \mathbf{b} = \mathbf{a}^\top \mathbf{Q}_z^\top \mathbf{Q}_y \mathbf{b} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top$.
Entonces las funciones son: $f(z) = \mathbf{Q}_z(z)a; \quad g(x) = \mathbf{Q}_y(y)\mathbf{b}$.

Optimizacion con flujos de partículas

Dados a y b terminamos en un problema de minimizacion dado por

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

Problema de optimizacion con mapas que mueven **todas** las particulas

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

esto lo puedo ver como un descenso de gradientes.

$$\nabla_y A^{hl}\Big|_{y_i^n} = Q_z^h(z_i) \sum_j \nabla G^j(y)\Big|_{y_i} B_{jl}^y$$