$$\min_{\theta} \; E_{P_{trg}(x)\,P(y|x)}\left[(y - f_\theta(x))^2 \right]$$

$$\Updownarrow$$

$$\max_{\theta} \; E_{P_{trg}(x)\,P(y|x)}\left[\log P_\theta(y|x) \right] \qquad P_\theta(y|x) = \mathcal{N}(y; f_\theta(x), \sigma^2)$$

$$\text{logloss}_{P_{trg}(x)}\left(P(y|x), \hat{P}(y|x) \right) = E_{P_{trg}(x)\,P(y|x)}\left[-\log \hat{P}(y|x) \right]$$

$$\text{rel-loss}_{P_{trg}(x)}\left(P(y|x), \hat{P}(y|x), P_0(y|x) \right) = E_{P_{trg}(x)\,P(y|x)}\left[-\log \frac{\hat{P}(y|x)}{P_0(y|x)} \right]$$

$$\Gamma = \left\{ P(y|x) \;\middle|\; E_{P_{src}(x)\,P(y|x)}\left[\Phi(x,y) \right] = c \right\}$$

- Robust Bias-Aware Regression

$$\min_{\hat{P}(y|x) \in \Delta} \; \max_{P(y|x) \in \Gamma \cap \Delta} \; E_{P_{trg}(x)\,P(y|x)}\left[-\log \frac{\hat{P}(y|x)}{P_0(y|x)} \right]$$

$$-\int_x \int_y P_{trg}(x)\, P(y|x) \log \frac{\hat{P}(y|x)}{P_0(y|x)}\, dy\, dx \;, \quad \text{Convex in } \hat{P}, \quad \text{Concave (affine) in } P$$

$$\Downarrow$$

$$\max_{P(y|x) \in \Gamma \cap \Delta} \; \min_{\hat{P}(y|x) \in \Delta} \; -\int_x \int_y P_{trg}(x)\, P(y|x) \log \frac{\hat{P}(y|x)}{P_0(y|x)}\, dy\, dx$$

Inner minimization:

$$\begin{cases} \min_{\hat{P}} \; -\int_x \int_y P_{trg}(x)\, P(y|x) \log \dfrac{\hat{P}(y|x)}{P_0(y|x)}\, dy\, dx \\ \text{s.t.} \; \int_y \hat{P}(y|x)\, dy = 1 \quad \forall x \\ \qquad \Downarrow \\ \min_{\hat{P}(y|x)} \; -\int_y P(y|x) \log \dfrac{\hat{P}(y|x)}{P_0(y|x)}\, dy \\ \text{s.t.} \; \int_y \hat{P}(y|x)\, dy = 1 \end{cases}$$

$$\mathcal{L} = -\int_y p(y|x) \log \frac{\hat{p}(y|x)}{p_0(y|x)} dy + \nu\left(\int_y \hat{p}(y|x)\, dy - 1 \right)$$

$$\nabla_{\hat{p}(y|x)} \mathcal{L} = -\frac{p(y|x)}{\hat{p}(y|x)} + \nu :=0 \qquad \hat{p}(y|x) = \frac{p(y|x)}{\nu}$$

$$g(\nu) = -\int_y p(y|x) \log \frac{p(y|x)}{\nu\, p_0(y|x)} dy + 1 - \nu$$

$$= \int_y p(y|x) \left[\log \nu\, p_0(y|x) - \log p(y|x) \right] dy + 1 - \nu$$

$$\nabla_\nu g = \int_y \frac{p(y|x)}{\nu} dy - 1 := 0 \qquad \nu^* = 1$$

$$\hat{p}^*(y|x) = p(y|x) / \nu^* = p(y|x)$$

$$\min_{p(y|x)} \int_x \int_y p_{tg}(x)\, p(y|x) \log \frac{p(y|x)}{p_0(y|x)} dy\, dx \qquad \text{convex in } p \quad (x \log x \text{ form})$$

$$\text{s.t.} \quad \int_y p(y|x)\, dy = 1 \quad \forall x \qquad \theta(x)$$

$$\int_x \int_y p_{src}(x)\, p(y|x)\, \Phi(x,y)\, dy\, dx = C \qquad \theta$$

$$\mathcal{L} = \int_x \int_y p_{tg}(x)\, p(y|x) \log \frac{p(y|x)}{p_0(y|x)} dy\, dx$$

$$+ \int_x \theta(x) \left[\int_y p(y|x)\, dy - 1 \right] dx$$

$$+ \theta^T \left(\int_x \int_y p_{src}(x)\, p(y|x)\, \Phi(x,y)\, dy\, dx - C \right)$$

$$\nabla_{p(y|x)} \mathcal{L} = p_{tg}(x) \left[1 + \log p(y|x) - \log p_0(y|x) \right] + \theta(x) + p_{src}(x)\, \theta^T \Phi(x,y) := 0$$

$$p(y|x) = \exp\left[-\frac{\theta(x)}{p_{tg}(x)} - \frac{p_{src}(x)}{p_{tg}(x)} \theta^T \Phi(x,y) - 1 + \log p_0(y|x) \right]$$

$$p_\theta(y|x) \propto \exp\left[-\frac{p_{src}(x)}{p_{tg}(x)} \theta^T \Phi(x,y) \right] \cdot p_0(y|x)$$

• Density Estimation

$$p_\theta(y|x) \propto \exp\left[-\frac{p_{src}(x)}{p_{trg}(x)}\theta^T\Phi(x,y)\right] p_0(y|x)$$

if $p_0(y|x) \sim \mathcal{N}(u_0, \sigma_0^2)$ is a Gaussian, $\theta^T\Phi(x,y) = [y, x, 1] M \begin{bmatrix} y\\ x\\ 1\end{bmatrix}$

then $p_\theta(y|x)$ is also a Gaussian

$$\theta = M = \begin{bmatrix} M_{11} & M_{12}\\ M_{21} & M_{22}\end{bmatrix}$$

$$\Phi(x,y) = \begin{bmatrix} y\\ x\\ 1\end{bmatrix}[y\ x\ 1]$$

$$p_\theta(y|x) \propto \exp\left[-\frac{p_{src}(x)}{p_{trg}(x)}[y\ x\ 1] M \begin{bmatrix} y\\ x\\ 1\end{bmatrix}\right] p_0(y|x)$$

$$\propto \exp\left\{-\frac{(y-u_0)^2}{2\sigma_0^2} - \frac{p_{src}(x)}{p_{trg}(x)}[y\ x\ 1] M \begin{bmatrix} y\\ x\\ 1\end{bmatrix}\right\}$$

$$\propto \exp\left\{-\frac{(y-u_0)^2}{2\sigma_0^2} - \frac{p_{src}(x)}{p_{trg}(x)}\left[y^2 M_{11} + 2y M_{12}\begin{bmatrix} x\\ 1\end{bmatrix}\right]\right\}$$

$$-\frac{(y-u_0)^2}{2\sigma_0^2} - \frac{p_{src}(x)}{p_{trg}(x)}\left[M_{11}y^2 + 2y M_{12}\begin{bmatrix} x\\ 1\end{bmatrix}\right]$$

$$= -\frac{1}{2\sigma_0^2}(y^2 - 2u_0 y + u_0^2) - M_{11}\frac{p_{src}(x)}{p_{trg}(x)}y^2 - 2M_{12}\begin{bmatrix} x\\ 1\end{bmatrix}\frac{p_{src}(x)}{p_{trg}(x)}y$$

$$\Rightarrow y^2\left(-\frac{1}{2\sigma_0^2} - M_{11}\frac{p_{src}(x)}{p_{trg}(x)}\right) + \left(2M_{12}\begin{bmatrix} x\\ 1\end{bmatrix}\frac{p_{src}(x)}{p_{trg}(x)} - \frac{u_0}{\sigma_0^2}\right)y$$

$$ay^2 + by \Rightarrow a\left(y^2 + \frac{b}{a}y\right) \Rightarrow -\frac{1}{2\cdot\frac{-1}{2a}}\left(y + \frac{b}{2a}\right)^2$$

$$\sigma^2 = -\frac{1}{2a} = \left(\frac{1}{\sigma_0^2} + 2M_{11}\frac{p_{src}(x)}{p_{trg}(x)}\right)^{-1}$$

$$u = \frac{b}{2a} = \left(\frac{1}{\sigma_0^2} + 2M_{11}\frac{p_{src}(x)}{p_{trg}(x)}\right)^{-1}\left(-2M_{12}\begin{bmatrix} x\\ 1\end{bmatrix}\frac{p_{src}(x)}{p_{trg}(x)} + \frac{u_0}{\sigma_0^2}\right)$$

$$p_\theta(y|x) \sim \mathcal{N}(u(x,\theta), \sigma^2(x,\theta))$$

$$u(x,\theta) = \left(\frac{1}{\sigma_0^2} + 2M_{11}\frac{p_{src}(x)}{p_{trg}(x)}\right)^{-1}\left(-2M_{12}\begin{bmatrix} x\\ 1\end{bmatrix}\frac{p_{src}(x)}{p_{trg}(x)} + \frac{u_0}{\sigma_0^2}\right)$$

$$\sigma^2(x,\theta) = \left(\frac{1}{\sigma_0^2} + 2M_{11}\frac{p_{src}(x)}{p_{trg}(x)}\right)^{-1}$$

- Gradient Method for Dual

$$\mathcal{L} = \int_x \int_y P_{trg}(x)\, P(y|x) \log \frac{P(y|x)}{P_0(y|x)}\, dy\, dx$$
$$+ \int_x \theta(x) \left[\int_y P(y|x)\, dy - 1 \right] dx$$
$$+ \theta^T \left(\int_x \int_y P_{src}(x)\, P(y|x)\, \Phi(x,y)\, dy\, dx - C \right)$$

$$= \int_x \int_y P_{trg}(x)\, P(y|x) \log \frac{P(y|x)}{P_0(y|x)}\, dy\, dx$$
$$+ \left\langle \mu, \int_x \int_y P_{src}(x)\, P(y|x) \begin{bmatrix} y \\ x \\ 1 \end{bmatrix} \begin{bmatrix} y & x & 1 \end{bmatrix} dy\, dx - C \right\rangle$$

evaluate primal optimal at dual

$$P_\mu(y|x) = \mathcal{N}\left(u(x,\theta), \sigma^2(x,\theta)\right)$$

$$u(x,\theta) = \left(\frac{1}{\sigma_0^2} + 2\mu_{11} \frac{P_{src}(x)}{P_{trg}(x)} \right)^{-1} \left(-2\mu_{12} \begin{bmatrix} x \\ 1 \end{bmatrix} \frac{P_{src}(x)}{P_{trg}(x)} + \frac{u_0}{\sigma_0^2} \right)$$

$$\sigma^2(x,\theta) = \left(\frac{1}{\sigma_0^2} + 2\mu_{11} \frac{P_{src}(x)}{P_{trg}(x)} \right)^{-1}$$

apply dual gradient ascent

$$\mu := \mu + \alpha \Bigg[\underbrace{\int_x \int_y P_{src}(x)\, P_\theta(y|x) \begin{bmatrix} y \\ x \\ 1 \end{bmatrix} \begin{bmatrix} y & x & 1 \end{bmatrix} dy\, dx}_{\text{estimated by sampling}} - C \Bigg] \underbrace{-\lambda\mu}_{\text{regularization}}$$

$$E[Y^2] = E[Y]^2 + \mathrm{var}(Y)$$