1 Priklad - Nejmensi ctverce

Zakladni rovnice kterou chceme naprogramovat.

$$\min_{\theta} -\mathbb{E}_{p_{\text{data}}(x,y)} \left[\alpha \log q_\theta\left(y|x\right) + (1-\alpha)\log q_\theta\left(x|y\right)\right] \tag{1}$$

$$\approx \min_{\theta} -\mathbb{E}_{p_{\text{data}}(x,y)} \left[\alpha \log \frac{\exp\left(f_\theta\left(x\right)\left[y\right]\right)}{\sum_{y'}\exp\left(f_\theta\left(x\right)\left[y'\right]\right)} + (1-\alpha)\log \frac{\exp\left(f_\theta\left(x\right)\left[y\right]\right)}{\sum_{i=1}^{K}\exp\left(f_\theta\left(x_i\right)\left[y\right]\right)}\right]. \tag{2}$$

Dle znaceni v clanku **Joint Energy models** plati

$$p(x,y) = \frac{\exp\left(f_\theta(x)[y]\right)}{Z(\theta)} \tag{3}$$

kde

$$f_\theta(x)[y] = -E_\theta(x,y). \tag{4}$$

Jak to premostit na nejmensi ctverce? Vime, ze pro distribuce plati

$$p(x,y) = p(y,x) = p(y|x)\cdot p(x) \tag{5}$$

$$p(y|x) = \mathcal{N}\left(\theta_0 + \theta_1 x, \sigma^2\right), \tag{6}$$

pripadne pro vyssi rad $\theta_0 + \theta_1 x + \dots \theta_k x^k$. Distribuci $p(x)$ muzeme zvolit libovolne. Napriklad pro jednoduchost

$$p(x) = \mathcal{N}\left(0, \sigma^2\tau^2\right), \tag{7}$$

pricemz uvazujeme σ za zname a τ volime dle potreby. Z cehoz plyne ze

$$p(x,y) = \mathcal{N}\left(0,\sigma^2\tau^2\right)\cdot\mathcal{N}\left(\theta_0+\theta_1 x,\sigma^2\right) = \frac{1}{2\pi\sigma\tau}\exp\left(-\frac{(y-\theta_0-\theta_1 x)^2}{2\sigma^2} - \frac{x^2}{2\sigma^2\tau^2}\right) \tag{8}$$

a tudiz

$$f_\theta(x)[y] = -\frac{(y-\theta_0-\theta_1 x)^2}{2\sigma^2} - \frac{x^2}{2\sigma^2\tau^2}. \tag{9}$$

Timto mame definovany nas model. Je potreba si vsak uvedomit, ze prvni diskrimantivni clen v rovnici (2) je standardni softmax slouzici pro klasifikaci. My v tomto okamziku nechceme klasifikovat, ale nalezt nejlepsi proklad danymi body. Proto diskrimanativni clen nahradime drouhou mocninou L_2 normy rozdilu $y - \theta_0 - \theta_1 x$. Tedy

$$\text{SSE} = \sum_{i=1}^{N}(y_i - \theta_0 - \theta_1 x_i)^2 \tag{10}$$

Nyni dosadme (9) a (10) do (2) a pouzijeme pravidla pro pocitani s logaritmy.

$$\min_\theta \left\{\alpha\ \text{SSE} - \mathbb{E}_{p_{\text{data}}(x,y)}\left[(1-\alpha)\log q_\theta\left(x|y\right)\right]\right\} = \tag{11}$$

$$\min_\theta \left\{\alpha\ \text{SSE} - \mathbb{E}_{p_{\text{data}}(x,y)}\left[(1-\alpha)\log\frac{\exp\left(f_\theta\left(x\right)\left[y\right]\right)}{\sum_x \exp\left(f_\theta\left(x_i\right)\left[y\right]\right)}\right]\right\} = \tag{12}$$

$$\min_\theta \left\{\alpha\ \text{SSE} - \mathbb{E}_{p_{\text{data}}(x,y)}\left[(1-\alpha)\log\frac{\exp\left(-\frac{(y-\theta_0-\theta_1 x)^2}{2\sigma^2} - \frac{x^2}{2\sigma^2\tau^2}\right)}{\sum_x \exp\left(-\frac{(y-\theta_0-\theta_1 x)^2}{2\sigma^2} - \frac{x^2}{2\sigma^2\tau^2}\right)}\right]\right\} \tag{13}$$

Pricemz generativni clen $q_\theta\left(x|y\right)$ se da zjednodusit do formy

$$\log q_\theta\left(x|y\right) = \left(-\frac{(y-\theta_0-\theta_1 x)^2}{2\sigma^2} - \frac{x^2}{2\sigma^2\tau^2}\right) - \log\sum_x \exp\left(-\frac{(y-\theta_0-\theta_1 x)^2}{2\sigma^2} - \frac{x^2}{2\sigma^2\tau^2}\right) \tag{14}$$