Manifold Learning via Conditional Entropy Minimization

Shashank Singh
sss1@andrew.cmu.edu

1 Introduction

Given samples $x_1, \dots, x_n \sim p$, where $p : \mathbb{R}^d \to \mathbb{R}_+$ is an unknown probability density, we are interested in finding a continuous surjection $f : \mathbb{R}^d \to [0,1]\}$ so as to minimize the ratio of the conditional entropy of $f(X)$ given X to the entropy of $f(X)$:

$$\begin{aligned} H(Y|X) &= -\int_{\mathcal{X}} p_X(x) \int_{\mathcal{Y}} p_{Y|X}(y|x) \log p_{Y|X}(y|x)\, dy\, dx \\ &= -\int_{\mathcal{X}\times\mathcal{Y}} p_{Y,X}(y,x) \log \frac{p_{Y,X}(y,x)}{p_X(x)}\, d(x,y) = -\mathbb{E}_{X,Y}\left[\log \frac{p_{Y,X}(y,x)}{p_X(x)}\right]. \end{aligned}$$

Hence, given $y_1, \dots, y_n$ and an estimator $\hat{p}$ for p, a reasonable estimator $\hat{H}(Y|X)$ for $H(Y|X)$ might be

$$\hat{H}(Y|X) = -\frac{1}{n}\sum_{i=1}^n \log\left(\frac{\hat{p}_{X,Y}(x_i,y_i)}{\hat{p}_X(x_i)}\right).$$

Note that

$$\begin{aligned} \frac{d}{dy_i}\hat{H}(Y|X) &= -\frac{1}{n}\sum_{i=1}^n \frac{d}{dy_i}\log\left(\frac{\hat{p}_{X,Y}(x_i,y_i)}{\hat{p}_X(x_i)}\right) \\ &= -\frac{1}{n}\sum_{i=1}^n \frac{d}{dy_i}\log\left(\hat{p}_{X,Y}(x_i,y_i)\right) = -\frac{1}{n}\sum_{i=1}^n \frac{\frac{d}{dy_i}\hat{p}_{X,Y}(x_i,y_i)}{\hat{p}_{X,Y}(x_i,y_i)}. \end{aligned}$$

Hence, if we use a kernel density estimate $\hat{p}$ of p with bandwidth h and symmetric kernel K_h, then $\forall k \in [n]$,

$$\begin{aligned} \frac{d}{dy_k}\hat{H}(Y|X) &= -\frac{1}{n}\sum_{i=1}^n \frac{\sum_{j=1}^n \frac{d}{dy_k}K_h((x_i,y_i),(x_j,y_j))}{\sum_{j=1}^n K_h((x_i,y_i),(x_j,y_j))} \\ &= -\frac{1}{n}\sum_{i=1}^n \frac{\sum_{j=1}^n K_h(x_i,x_j)\frac{d}{dy_k}K_h(y_i,y_j)}{\sum_{j=1}^n K_h(x_i,x_j)K_h(y_i,y_j)} \\ &= -\frac{1}{n}\left(\frac{\sum_{j=1}^n K_h(x_k,x_j)\frac{d}{dy_k}K_h(y_k,y_j)}{\sum_{j=1}^n K_h(x_k,x_j)K_h(y_k,y_j)} - \sum_{i=1}^n \frac{K_h(x_i,x_k)\frac{d}{dy_k}K_h(y_k,y_i)}{\sum_{j=1}^n K_h(x_i,x_j)K_h(y_i,y_j)}\right) \end{aligned}$$

Similarly,

$$\hat{H}(Y) = -\int_{\mathcal{Y}} p_Y(y)\log p_Y(y)\, dy = -\mathbb{E}_Y\left[\log p_Y(y)\right].$$

Hence, given $y_1, \dots, y_n$ and an estimator $\hat{p}_Y$ for p_Y, a reasonable estimator $\hat{H}(Y)$ for $H(Y)$ might be

$$\hat{H}(Y) = -\frac{1}{n}\sum_{i=1}^n \log\left(\hat{p}_Y(y_i)\right).$$

$$
\begin{aligned}
\frac{d}{dy_i}\hat{H}(y) = -\frac{1}{n}\sum_{i=1}^{n}\frac{d}{dy_i}\log\left(\hat{p}_Y(y_i)\right) &= -\frac{1}{n}\sum_{i=1}^{n}\frac{\frac{d}{dy_i}\frac{1}{nh}\sum_{j=1}^{n}K_h(y_i,y_j)}{\frac{1}{nh}\sum_{j=1}^{n}K_h(y_i,y_j)} \\
&= -\frac{1}{n}\sum_{i=1}^{n}\frac{\sum_{j=1}^{n}\frac{d}{dy_i}K_h(y_i,y_j)}{\sum_{j=1}^{n}K_h(y_i,y_j)}
\end{aligned}
$$