Machine Learning and Physics, Sheet 02

Jacob Waßelewsky, Johannes Schmidt, Dominik Geng

1 Exercise 3: Mean-Shift

Gradient ascent on the KDE with the Epanechnikov kernel corresponds to the update step

$$x_j^{t+1} = x_j^t + \alpha_j^t \frac{2}{n} \sum_{i:||x_i - x_j^t||<1} (x_i - x_j^t) \tag{1}$$

with $\mathbf{x} \in \mathbb{R}^n$.
We define I_j as the set of x_i that lie inside a distance of 1 from x_j,

$$I_j := \{x_i : ||x_i - x_j||< 1\}. \tag{2}$$

Setting α_j now proportional to the fraction of number of elements in I_j out of $\mathbf{x}$,

$$\alpha_j = \frac{n}{2|I_j|} \tag{3}$$

we obtain

$$\begin{aligned} x_j^{t+1} &= x_j + \frac{n}{2|I_j|} \frac{2}{n} \sum_{i:||x_i - x_j^t||<1} (x_i - x_j^t) \\ &= x_j + \frac{1}{|I_j|} \sum_{i:||x_i - x_j^t||<1} (x_i - x_j^t). \end{aligned} \tag{4}$$

This implies, that we shift each data point to its local mean.

2 Exercise 4: K-Means

We aim to cluster a dataset $\mathbf{X} \in \mathbb{R}^{p\times N}$ into K clusters, by choosing cluster centers $\mathbf{C} \in \mathbb{R}^{p\times K}$ and cluster memberships $\mathbf{M} \in [0,1]^{K\times N}$ with $\sum_k M_{kn} = 1$, such that

$$E(\mathbf{C}, \mathbf{M}; K) = ||\mathbf{X} - \mathbf{CM}||^2 = \sum_{n=1}^{N} \sum_{k=1}^{K} m_{kn} ||\mathbf{x}_n - \mathbf{c}_k||^2 \tag{5}$$

is minimized. We'll divide the optimization into two steps.
1. **Assignment**: m_{kn}
Here, each data point $\mathbf{x_n}$ is assigned to its nearest mean. In order to minimize E w.r.t. m_{kn} consider the class assignment $\hat{k}_n$

$$\hat{k}_n = \arg\min_k ||\mathbf{x}_n - \mathbf{c}_k||^2 \tag{6}$$

Then, for a "hard" clustering, choose

$$m_{kn} = \begin{cases} 1 & \text{if } k = \hat{k}_n \\ 0 & \text{if } k \neq \hat{k}_n \end{cases} \tag{7}$$

which satisfies

$$\sum_{k}^{K} m_{kn} = 1.$$

2. **Update**: $\mathbf{c}_k$
In this step we update the means $\mathbf{c}_k$ to the sample means of their assigned cluster members. Rewriting the loss as

$$\sum_{n=1}^{N}\sum_{k=1}^{K} m_{kn}||\mathbf{x}_n - \mathbf{c}_k||^2 = \sum_{n}^{N}\left(m_{1n}||\mathbf{x}_n - \mathbf{c}_1||^2 + \ \dots \ + m_{Kn}||\mathbf{x}_n - \mathbf{c}_K||^2\right),$$

we see that in order to minimize E, we have to take

$$\hat{\mathbf{c}}_k = \arg\min_{\mathbf{c}_k} \sum_{n}^{N} m_{kn}||\mathbf{x}_n - \mathbf{c}_k||^2. \tag{8}$$

Therefore we set

$$\begin{aligned}
0 \overset{!}{=} \frac{\partial}{\partial \mathbf{c}_k} \sum_{n}^{N} m_{kn}||\mathbf{x}_n - \mathbf{c}_k||^2 &= \sum_{n}^{N} m_{kn} 2(\mathbf{c}_k - \mathbf{x}_n) \\
\Rightarrow \mathbf{c}_k \sum_{n}^{N} m_{kn} = \sum_{n}^{N} m_{kn} x_n & \\
\Rightarrow \hat{\mathbf{c}}_k = \frac{\sum_{n}^{N} m_{kn} x_n}{\sum_{n}^{N} m_{kn}} & \qquad (9)
\end{aligned}$$

where we have used that

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{c}_k} m_{kn}||\mathbf{x}_n - \mathbf{c}_k||^2 &= 2||\mathbf{x}_n - \mathbf{c}_k||\frac{\partial}{\partial \mathbf{c}_k}||\mathbf{x}_n - \mathbf{c}_k|| \\
&= 2||\mathbf{x}_n - \mathbf{c}_k||\frac{\mathbf{c}_k - \mathbf{x}_n}{||\mathbf{x}_n - \mathbf{c}_k||} \\
&= 2(\mathbf{c}_k - \mathbf{x}_n) \qquad (10)
\end{aligned}$$