867002105_HW6

Charles Dotson

October 21, 2022

1 K-means on the real line. Given the unlabeled training data $x^{(1)} = 0, x^{(2)} = 1, x^{(3)} = -2$, find the two centers μ_1 and μ_2 on the real line and the corresponding assignment r_{ik} of the training data to the two clusters, so that the distortion measure $J(\mu_1, \mu_2) = \sum_{i=1}^{3} \sum_{k=1}^{2} r_{ik}(x^{(i)} - \mu_k)^2$ is minimized. What's the smallest J value?

We start by noticing that in K-means, that one point cannot be in two clusters. Thus with the trainning data, we can right away put $x^{(1)}, x^{(2)}$ in the $k = 1$ cluster. We then move onto deriving the distortion measure J.

$$J(\mu_1, \mu_2) = \sum_{i=1}^{3} \sum_{k=1}^{2} r_{ik}(x^{(i)} - \mu_k)^2$$

$$= \sum_{i=1}^{3} \sum_{k=1}^{2} r_{ik}((x^{(i)})^2 - 2x^{(i)}\mu_k - \mu_k^2)$$

$$= \sum_{i=1}^{3} [r_{i1}((x^{(i)})^2 - 2x^{(i)}\mu_1 - \mu_1^2) + r_{i2}((x^{(i)})^2 - 2x^{(i)}\mu_2 - \mu_2^2)]$$

After expanding the sum and plugging on our trainging data as well as our chosen r_{ik} we have already chosen. We arrive at.

$$J(\mu_1, \mu_2) = 2\mu_1^2 - 2\mu_1 + 5 + 4\mu_2 + \mu_2^2$$

To find the minimum of this surface we start by finding the gradient of J w.r.t μ_k.

$$\nabla J = \begin{bmatrix} 4\mu_1 - 2 \\ 2\mu_2 + 4 \end{bmatrix}$$

To find the minimum of this fucntion both of these entries in the gradient vector will be set to zero and solved for μ_k and thus we have. $\mu_1 = \frac{1}{2}, \mu_2 = -2$. Thus minimizing the distortion function $J(1/2, -2) = 1/2$.

2 The input vector x is generated from a Gaussian $N(x|\mu_k, \sum_k)$ with probability ϕ_k, so that $Pr(x) = \sum_{k=1}^{K} \phi_k N(x|\mu_k, \sum_k)$. Let the one-hot vector z $\in 0, 1^k, \sum_k z_k = 1$ indicate which cluster x belongs to, and $Pr(z_k = 1) = \phi_k$. Prove that

$$\sum_{k=1}^{K} \phi_k N(x|\mu_k, \Sigma_K) = \sum_z \prod_{k=1}^{k} \phi_k^{z_k} \prod_{k=1}^{k} N(x|\mu_k, \Sigma_k)^{z_k}$$

$$Pr(x) = \sum_z Pr(z)Pr(x|z)$$

$$= \sum_z \prod_{k=1}^{K} \phi_k^{z_k} \prod_{k=1}^{K} N(x|\mu_k, \Sigma_k)^{z_k}$$

$$= \sum_z \prod_{k=1}^{K} \phi_k^{z_k} N(x|\mu_k, \Sigma_k)^{z_k}$$

Since we can see that only one-hot vector will have 1 and the other entries will be 0. We can see that.

$$= \sum_z \phi_k N(x|\mu_k, \Sigma_k) \quad \text{for which ever k where z=1}$$

Thus, instead of summing over all the hot vectors in the simulation, it is equivalent to saying summing over all of the k-th Gaussians. Thus,

$$= \sum_{k=1}^{K} \phi_k N(x|\mu_k, \Sigma_k)$$

3 Question 5

Excuse me not putting the prompt as it is quite large.

ran out of time to latex the proof but will email picture of what I have written down. apologiezes