Cover page for answers.pdf
CSE512 Fall 2018 - Machine Learning - Homework 7

Your Name: Sriram Reddy Kalluri

Solar ID: 111878857

NetID email address: skalluri@cs.stonybrook.edu and sriram.kalluri@stonybrook.edu

Names of people whom you discussed the homework with:

1) Expectation - Maximization.

During the expectation step, we will be finding the probabilities $P(z^i | x^i, \theta^{old})$ for all values z^i.

As stated in the problem, they have given the probabilities which are the output of E step.

$$R = \begin{pmatrix} 1 & 0 \\ 0.3 & 0.7 \\ 0 & 1 \end{pmatrix} \Rightarrow \quad P(z^i | x^i, \theta^{old}) = R_{i, z^i}$$

where the probability of observation x_i belonging to the cluster z^i.

During the maximization step, the likelihood function which we are trying to optimize.

$$\theta^{new} = \underset{\theta}{\operatorname{argmax}}\, Q(\theta, \theta^{old}).$$

where $Q(\theta, \theta^{old}) = \sum_{i=1}^{n} \sum_{z_i} P(z^i | x^i, \theta^{old}) \log(P(x^i, z^i | \theta))$

$$= \sum_{i=1}^{3} \sum_{k=z_i}^{2} P(z^i | x^i, \theta^{old}) \log(P(x^i, z^i | \theta))$$

where $P(x^i, z^i | \theta)$ is the probability of observation x^i given the parameters of the gaussian θ i.e μ_i, σ_i.

and $P(z^i | x^i, \theta^{old})$ is the probability of ith cluster.

where θ^{old} is the set of all parameters in model $\mu_1, \mu_2, \sigma_1, \sigma_2, \pi_1, \pi_2$.

2) . As discussed in class .

We know that π_i is the probability of i^{th} cluster .

$$\pi_i = \frac{1}{N} \sum_{j=1}^{N} P(z^i | x^j, \theta) .$$

$$\text{Given} \quad R = \begin{bmatrix} 1 & 0 \\ 0.3 & 0.7 \\ 0 & 1 \end{bmatrix}$$

$$\pi_1 = \frac{1}{N} \sum_{i=1}^{3} R_{i1} = \frac{1.3}{3} = 0.433$$

$$\pi_2 = \frac{1}{N} \sum_{i=1}^{3} R_{i2} = \frac{1.7}{3} = 0.566$$

---

3) As discussed in class,

$$\mu_k = \frac{\sum_{i=1}^{3} R_{ik} x^i}{\sum_{i=1}^{3} R_{ik}}$$

$x = [1 . 10 \quad 20]$

$$\mu_1 = \frac{1 \times 1 + 0.3 \times 10}{1.3} = \frac{4}{1.3} = 3.077$$

$$\mu_2 = \frac{20 \times 1 + 10 \times 0.7}{1.7} = \frac{27}{1.7} = 15.882 .$$

4). $\sigma_k^2 = \frac{\sum_{i=1}^{3} R_{ik} x^{i\,2}}{\sum_{i=1} R_{ik}} - \mu_k^2.$

Applying this formula,

$$\sigma_1^2 = \frac{1\times1^2 + 0.3\times10^2}{1.3} - (3.077)^2.$$

$$= \frac{31}{1.3} - (3.077)^2 = 14.37822.$$

$$\sigma_1 = \sqrt{14.37822} = 3.791.$$

$$\sigma_2^2 = \frac{(20)^2\times1 + 10^2\times0.7}{1.7} - (15.882)^2$$

$$= \frac{470}{1.7} - (15.882)^2 = 276.47 - 252.1744$$
$$= 24.2961$$

$$\sigma_2 = 4.929.$$

E - step:

1). $R_{ij} = P(z^j | x^i; \theta) = \frac{P(x^i | \theta, z^j).P(z^j)}{\sum_{z^j} P(x^i|\theta, z^j).P(z^j)}$

This is based on bayesian rule.

where. $P(x^i|\theta, z^j) = \frac{1}{\sqrt{2\pi}\,\sigma_j} \exp\left(-\frac{(x^i-\mu_j)^2}{2\sigma_j^2}\right).$

$P(z^j) = \pi_j$ calculated in expectation step.

where. x^i — current observation.

μ_j, σ_j — mean, variance calculated in expectation step.

2) 'Applying' the formula.

$$\gamma_{11} = \frac{\frac{\pi_1}{\sigma_1\sqrt{2\pi}} \exp\left(-\frac{(x^1-\mu_1)^2}{2\sigma_1^2}\right)}{\frac{\pi_1}{\sigma_1\sqrt{2\pi}} \exp\left(-\frac{(x^1-\mu_1)^2}{2\sigma_1^2}\right) + \frac{\pi_2}{\sigma_2\sqrt{2\pi}} \exp\left(-\frac{(x^1-\mu_2)^2}{2\sigma_2^2}\right)}$$

$$= \frac{\frac{0.433}{3.791} \exp\left(-\frac{(1-3.07)^2}{2\times14.378}\right)}{\left(\frac{0.433}{3.791}\right) \exp\left(-\frac{(1-3.07)^2}{2\times14.3778}\right) + \frac{0.566}{4.929} \exp\left(-\frac{(1-15.882)^2}{2\times24.2961}\right)}$$

$$= 0.987912.$$

$$\gamma_{12} = 1-\gamma_{11} = 0.01208.$$

$$\gamma_{21} = \frac{\left(\frac{0.433}{3.791}\right) \exp\left(-\frac{(10-3.07)^2}{2\times14.378}\right)}{\left(\frac{0.433}{3.791}\right) \times \exp\left(-\frac{(10-3.07)^2}{2\times14.378}\right) + \frac{0.566}{4.929} \times \exp\left(-\frac{(10-15.882)^2}{2\times24.2961}\right)}$$

$$= 0.2761$$

$$\gamma_{22} = 1-\gamma_{21} = 0.72381.$$

$$\gamma_{31} = \frac{\left(\frac{0.433}{3.791}\right) \times \exp\left(-\frac{(20-3.07)^2}{2\times14.378}\right)}{\left(\frac{0.433}{3.791}\right) \times \exp\left(-\frac{(20-3.07)^2}{2\times14.378}\right) + \frac{0.566}{4.929} \times \exp\left(-\frac{(20-\cancel{3.07}\,15.882)^2}{2\times24.2961}\right)}$$

$$= 6.611\times10^{-5}.$$

$$\gamma_{32} = 0.99993388.$$

$$R := \begin{bmatrix} 0.98791 & 0.01208 \\ 0.2761 & 0.72381 \\ 6.611\times10^{-5} & 0.99993 \end{bmatrix}$$

Please find my Kaggle rank and the accuracy below:

16	new	**Sriram Reddy Kalluri**	0.82400	5	now

I have trained five networks as given above. All are behaving in a similar way. Please find the network architecture which is giving the best validation accuracy.

1)Linear Layer(75,100)

2)Bi Directional LSTM Layer(hidden_size = 200, input_size = 100,num_layers=2,dropout=0.3, batch_first=True, bidirectional=True)

3)Bi Directional LSTM(hidden_size = 500, input_size = 400,num_layers=4,dropout=0.3, batch_first=True, bidirectional=True)

4)Linear Layer(1000, 10)

Initially I used SGD as optimizer. Then I changed to ADAM and RMSprop. But gave faster convergence results. I have used learning rate as 1e-3. It took 60 epochs to converge and the convergence criteria I used is if there is not much change for 5 epochs in training accuracy, it stops running the next epoch.

Report your Kaggle Performance here: 0.824 --- Rank 16