Rohan Mahendra Chaudhari

USCID : 6675-5653-85

1 Problem 1 - Optimization over the simplex

In this exercise you will prove two optimization results over the simplex that we used multiple times in the lectures. These results will also help you solve other problems in this homework. The K 1 dimensional simplex is simply the set of all distributions over K elements, denoted by $\Delta = \{\boldsymbol{q} \in R^K | q_k \geq 0, \forall k \text{ and } \sum_{k=1}^{K} q_k = 1\}$

1.1 Let $a_1, ..., a_K$ be K positive numbers. Prove that the solution of the following optimization problem

$$arg \max_{q \in \Delta} \sum_{k=1}^{K} a_k ln q_k$$

is q^* such that $q_k^* = \frac{a_k}{\sum_{k'} a_k'}$ (that is, $q_k^* \propto a_k$).

L$(\boldsymbol{q}, \lambda, \lambda_1, ..., \lambda_K) = \sum_{k=1}^{K} a_k ln q_k + \lambda(\sum_{k=1}^{K} q_k - 1) + \sum_{k=1}^{K} \lambda_k q_k$ for lagrangian multipliers $\lambda \neq 0$ and $\lambda_1, ..., \lambda_k \geq 0$. Now apply KKT conditions to find $\boldsymbol{q}^*$.

Solution

As per the stationary conditions,

For each K,

$\frac{a_k}{q_k^*} + \lambda + \lambda_k = 0$

$q_k^* = -\frac{a_k}{\lambda + \lambda_k} \neq 0$

As per the complementary slackness conditions, $\lambda_k q_k^* = 0$ and $\lambda_k = 0$

$\sum_{k=1}^{K} q_k^* = -\sum_{k=1}^{K} \frac{a_k}{\lambda} = 1$

$\lambda = -\sum_{k=1}^{K} a_k$

On substituting λ we get,

$q_k^* = -\frac{a_k}{\lambda}$

Which proves that $q_k^* \propto a_k$ with λ as a constant.

1.2 Let $b_1, ..., b_K$ be K real numbers and H be an entropy function. Prove that the solution of the following optimization problem

$$arg \max_{q \in \Delta} \sum_{k=1}^{K} a_k ln q_k$$

Solution

Lagrangian :
$L(\boldsymbol{q}, \lambda, \lambda_1, ..., \lambda_K) = \sum_{k=1}^{K}(q_k b_k - q_k ln q_k) + \lambda(\sum_{k=1}^{K} q_k - 1) + \sum_{k=1}^{K} \lambda_k q_k$
As per the stationary condition,
For each K,
$b_k - 1 - ln q_k + \lambda + \lambda_k = 0$
$q_k = e^{b_k - 1 + \lambda + \lambda_k} \propto e^{b_k + \lambda_k} \neq 0$
As per the solution for question 1.1 the complementary slackness implies $\lambda_k = 0$ and thus $q_k \propto e^{b_k}$

1.3 In the lecture we derived EM through a lower bound of the log-likelihood function. Specifically, on Slide 42 of Lecture 8, we find the tightest lower bound by solving

Solution

By using the result of Problem 1.2 with $b_k = \ln p(\boldsymbol{x}_n, z_n = k; \theta^{(t)})$
$q^*_{nk} \propto p(\boldsymbol{x}_n, z_n = k; \theta^{(t)})$
Can also be written as,
$q^*_{nk} = \frac{p(\boldsymbol{x}_n, z_n = k; \theta^{(t)})}{\sum_{k=1}^{K} p(\boldsymbol{x}_n, z_n = k; \theta^{(t)})} = \frac{p(\boldsymbol{x}_n, z_n = k; \theta^{(t)})}{p(\boldsymbol{x}_n; \theta^{(t)})} = p(z_n = k | \boldsymbol{x}_n; \theta^{(t)})$

2 Problem 2 - Gaussian Mixture Model

In the lecture we applied EM to learn Gaussian Mixture Models (GMMs) and showed the M-Step without a proof on Slide 48 of Lecture 8 . In this problem you will prove this for the simpler one-dimensional case. Specifically consider a one-dimensional GMM that has the following density function for x:

Solution

To find $\omega_1, ..., \omega_K$
Solve $\text{argmax}_{\omega \in \Delta} \sum_n \sum_k \gamma_{nk} \ln \omega_k$
According to the Hint given, Considering Problem 1.1 with $a_k = \sum_n \gamma_{nk}$,
$\omega_k = \frac{\sum_n \gamma_{nk}}{\sum_k \sum_n \gamma_{nk}} = \frac{\sum_n \gamma_{nk}}{\sum_n 1} = \frac{\sum_n \gamma_{nk}}{N}$
To find μ_k and σ_k, we solve for each k,
$\text{argmax}_{\mu_k, \sigma_k} \sum_n \gamma_{nk} \ln N(x_n | \mu_k, \sigma_k) = \text{argmax}_{\mu_k, \sigma_k} \sum_n \gamma_{nk} (ln \frac{1}{\sigma_k} - \frac{(x_n - \mu_k)^2}{2\sigma_k^2})$
Find the derivative with respect to μ_k and set it to 0,

$\frac{1}{\sigma_k^2}\sum_n \gamma_{nk}(x_n - \mu_k) = 0$
$\mu_k = \frac{1}{\sum_n \gamma_{nk}}\sum_n \gamma_{nk}x_n$
Find the derivative with respect to σ_k and set it to 0,
$\sum_n \gamma_{nk}(-\frac{1}{\sigma_k} + \frac{(x_n-\mu_k)^2}{\sigma_k^3}) = 0$
$\sigma_k^2 = \frac{1}{\sum_n \gamma_{nk}}\sum_n \gamma_{nk}(x_n - \mu_k)^2$
Thus, all the 3 values match the results mentioned in the question.

3 Problem 3 - EM

Consider the following probabilistic model to generate a non-negative integer x. First, draw a hidden binary variable z from a Bernoulli distribution with mean $\pi \in [0,1]$, that is, $p(z = 1, \pi) = \pi$ and $p(z = 0; \pi) = 1 - \pi$. If z=0, set x=0; otherwise, draw x from a Poisson distribution with parameter λ so that $p(x|z = 1; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$
Given N samples $x_1, ..., x_N$ generated independently in this way, follow the steps below to derive the EM algorithm for this model. (This is also a good example to see why finding the exact MLE is difficulty; you might try it yourself.)

3.1 Fixing the model parameters π and λ, for each sample n, find the posterior distribution of the corresponding hidden variable $z_n : p(z_n|x_n; \pi, \lambda)$. You will find it useful to consider the cases $x_n > 0$ and $x_n = 0$ separately. Given that z_n is binary, this means that you have to find the value of the following four quantities:

Solution
$\gamma_0^{'}$ = p($z_n = 0 \mid x_n > 0; \pi, \lambda) = 0$
$\gamma_1^{'}$ = p($z_n = 1 \mid x_n > 0; \pi, \lambda) = 1$
Here, as given in the question, $x_n = 0$ only when $z_n = 0$ thus when $x_n > 0$, z_n will always equal 1.
γ_0 = p($z_n = 0 \mid x_n = 0$;π, λ) $\propto$ p($x_n = 0 \mid z_n = 0$;π, λ)*p($z_n = 0$; π) = 1(1-π)
γ_1 = p($z_n = 1 \mid x_n = 0$; π, λ) $\propto$ p($x_n = 0 \mid z_n = 1$; π, λ)*p($z_n = 1$; π) = $e^{-\lambda}\pi$
Thus,
$\gamma_0 = \frac{1-\pi}{(1-\pi)+\pi e^{-\lambda}}$
$\gamma_1 = \frac{\pi e^{-\lambda}}{1-\pi+\pi e^{-\lambda}}$

3.2 **Suppose that we have computed $\gamma_0, \gamma_1, \gamma_0^{'}, \gamma_1^{'}$ from the previous value of π λ. Now, write down the expected complete likelihood function Q(π,λ) in terms of the data $x_1, ..., x_n$, the posteriors $\gamma_0, \gamma_1, \gamma_0^{'}, \gamma_1^{'}$, and the parameters π and λ (show intermediate steps). This completes the E-step.**

Solution

Q(π,λ) = $\sum_{x_n>0} \gamma_0^{'}$ ln $p(x_n > 0, z_n = 0; \pi, \lambda)$ + $\sum_{x_n>0} \gamma_1^{'}$ ln $p(x_n > 0, z_n = 1; \pi, \lambda)$ + $\sum_{x_n=0} \gamma_0$ ln $p(x_n = 0, z_n = 0; \pi, \lambda)$ + $\sum_{x_n=0} \gamma_1$ ln $p(x_n = 0, z_n = 1; \pi, \lambda)$

= 0 + $\sum_{x_n>0}$ 1∗ln $p(x_n > 0, z_n = 1; \pi, \lambda)$ + $\sum_{x_n=0} \gamma_0$ ln $p(x_n = 0, z_n = 0; \pi, \lambda)$ + $\sum_{x_n=0} \gamma_1$ ln $p(x_n = 0, z_n = 1; \pi, \lambda)$

= $\sum_{x_n>0}$ ln $(\pi \frac{\lambda^{x_n} e^{-\lambda}}{x_n!})$ + $\sum_{x_n=0} \gamma_0$ ln $(1 - \pi)$ + $\sum_{x_n=0} \gamma_1$ ln $(\pi e^{-\lambda})$

= $\sum_{x_n>0}$ (ln π + x_n ln λ - λ - ln(x_n!)) + $\sum_{x_n=0} \gamma_0$ ln $(1 - \pi)$ + $\sum_{x_n=0} \gamma_1$ (ln π - λ)

3.3 **Find the maximizer π^* and λ^* for the function Q(π,λ) from the previous question (show your derivation). You might find it convenient to use the notation $N_0 = |n : x_n = 0|$ (that is, the number of examples with value 0) in your solution. This completes the M-step.**

Solution

Q(π, λ) = $\sum_{x_n>0}$ (ln π + x_n ln λ - λ - ln(x_n!)) + $\sum_{x_n=0} \gamma_0$ ln $(1 - \pi)$ + $\sum_{x_n=0} \gamma_1$ (ln π - λ)

$\frac{\partial Q}{\partial \pi} = \sum_{x_n>0} \frac{1}{\pi} + 0 + 0 + 0 + \sum_{x_n=0} \gamma_0 \frac{1}{1-\pi} + \sum_{x_n=0} \gamma_1 \frac{1}{\lambda} = 0$

$\frac{N_{x>0}}{\pi} + \sum_{x_n=0} \gamma_0 \frac{-1}{1-\pi} + \frac{1}{\pi}$ - $\gamma_0 \frac{1}{\pi}$

Here, γ_1 = 1 - γ_0

$\frac{N_{x>0}}{\pi} + \frac{N_0}{\pi} + \sum_{x_n=0} \gamma_0 (\frac{-\pi-1+\pi}{\pi(1-\pi)})$

$\frac{N}{\pi} = \sum_{x_n=0} \gamma_0 \frac{1}{\pi(1-\pi)}$

$(1 - \pi) = \frac{N_0}{N} \gamma_0$

π^* = 1 - $\frac{N_0}{N} \gamma_0$

$\frac{\partial Q}{\partial \lambda} = \sum_{x_n>0} \frac{x_n}{\lambda}$ - 1 - $\sum_{x_n=0} \gamma_1$

= $\sum_{x_n>0} \frac{x_n}{\lambda}$ - $N_{x>0}$ - $N_0 \gamma_1$

We know that, $\gamma_1 = 1 - \gamma_0$

= $\sum_{x_n>0} \frac{x_n}{\lambda}$ - $N_{x>0}$ - N_0 + $N_0 \gamma_0$

= $\sum_{x_n>0} \frac{x_n}{\lambda}$ - N + $N_0 \gamma_0 = 0$

$\lambda^* = \frac{\sum_{x_n>0} x_n}{N - N_0 \gamma_0}$

3.4 Combining all the results, write down the EM update formulas for π^{new} and λ^{new}, using only the data $x_1, ..., x_n$ and the previous parameter values π^{old} and λ^{old} (do not use $\gamma_0, \gamma_1, \gamma_0^{'}, \gamma_1^{'}$).

Solution

$\pi^{new} = 1$ - $\frac{N_0}{N}$ $(\frac{1-\pi^{old}}{1-\pi^{old}+\pi^{old}e^{-\lambda^{old}}})$

$\lambda^{new} = \frac{\sum_{x_n>0} x_n}{N-N_0(\frac{1-\pi^{old}}{1-\pi^{old}+\pi^{old}e^{-\lambda^{old}}})}$