Quiz 7: Solutions

Consider the general discrete distribution $P(X = c) = \theta_c$ where $c \in \{1, ..., C\}$. Such a distribution must satisfy $\sum_{c=1}^{C} \theta_c = 1$ and $\theta_c \geq 0$ for all $1 \leq c \leq C$ since the distribution has to be normalized and satisfy non-negativity. A common penalized likelihood objective function for optimizing θ given a data set $\mathcal{D} = [x_1, ..., x_N]$is shown below where $\alpha > 0$ is a constant:

$$l(\mathcal{D}, \theta) = \sum_{n=1}^{N} \sum_{c=1}^{C} [x_n = c] \log(\theta_c) + \alpha \sum_{c=1}^{C} \log(\theta_c)$$

Due to the constraints $\sum_{c=1}^{C} \theta_c = 1$ and $\theta_c \geq 0$, this problem requires constrained optimization methods to solve. Use this information to answer the following two questions.

1 Question 1 (50 points)

As your answer to this question, derive the Lagrangian function for the constrained optimization problem described above. Briefly explain your answer.

Example Solution: The normalization constraint can be re-written in standard form as $\sum_{c=1}^{C} \theta_c - 1 = 0$. The non-negativity constraint is already given in standard form $\theta_c \geq 0$ for all $1 \leq c \leq C$. We will use λ as the Lagrange multiplier for the equality constraint, and γ_c as the Lagrange multipliers for the inequality constraints for class c. We this have:

$$L(\mathcal{D}, \theta, \lambda, \gamma) = \sum_{n=1}^{N} \sum_{c=1}^{C} [x_n = c] \log(\theta_c) + \alpha \sum_{c=1}^{C} \log(\theta_c) - \lambda \left(\sum_{c=1}^{C} \theta_c - 1 \right) - \sum_{c=1}^{C} \gamma_c \theta_C$$

Rubric:

- 10 points: correct objective function terms (first two terms)
- 15 points: correct normalization constraint term
- 15 points: correct non-negativity constraint term
- 10 points: correct explanation of solution
- -5 points: incorrect signs on Lagrange multipliers
- -5 points: missing index c on Lagrange multiplier for non-negativity constraint
- -5 points: sign error converting normalization constraint to canonical form
- -10 points: missing Lagrange multipliers

2 Question 2 (50 points)

As your answer to this question, use the Lagrangian to derive the optimal parameter estimates for the constrained optimization problem described above. Show your work.

Example Solution: Since the problem has both equality and inequality constraints, we have no solution method for the problem. Our only resort is to discard the inequality constraints and hope the equality constrained solution will also satisfy the inequality constraints. This will allow us to use the method of Lagrange multipliers:

$$\frac{\partial L(\mathcal{D}, \theta, \lambda, \gamma)}{\partial \theta_1} = \frac{\sum_{n=1}^{N}[x_n = 1]}{\theta_1} + \frac{\alpha}{\theta_1} - \lambda \tag{1}$$

$$\vdots$$

$$\frac{\partial L(\mathcal{D}, \theta, \lambda, \gamma)}{\partial \theta_C} = \frac{\sum_{n=1}^{N}[x_n = C]}{\theta_C} + \frac{\alpha}{\theta_C} - \lambda \tag{2}$$

$$\frac{\partial L(\mathcal{D}, \theta, \lambda, \gamma)}{\partial \lambda} = -(\sum_{c=1}^{C} \theta_c - 1) \tag{3}$$

We have a system of equations in $C + 1$ variables and our goal is to solve the system. Setting the last equation equal to 0, we recover the original constraint $\sum_{c=1}^{C} \theta_c = 1$. We deal with each of the gradient equations for the θ_c variables as follows:

$$\frac{\sum_{n=1}^{N}[x_n = c]}{\theta_c} + \frac{\alpha}{\theta_c} - \lambda = 0 \tag{4}$$

$$\sum_{n=1}^{N}[x_n = c] + \alpha = \lambda\theta_c \tag{5}$$

Now, we can eliminate the θ_c's from the above set of equations by adding all C equations together and using the normalization constraint:

$$\sum_{c=1}^{C}\left(\sum_{n=1}^{N}[x_n = c] + \alpha\right) = \lambda\sum_{c=1}^{C}\theta_c \tag{6}$$

$$\sum_{c=1}^{C}\sum_{n=1}^{N}[x_n = c] + C\alpha = \lambda(1) \tag{7}$$

$$N + C\alpha = \lambda \tag{8}$$

We can now solve Equation 5 for θ_c and substitute in the value of λ:

$$\lambda\theta_c = \sum_{n=1}^{N}[x_n = c] + \alpha \tag{9}$$

$$\theta_c^* = \frac{\sum_{n=1}^{N}[x_n = c] + \alpha}{\lambda} \tag{10}$$

$$= \frac{\sum_{n=1}^{N}[x_n = c] + \alpha}{N + C\alpha} \tag{11}$$

Finally, we need to check whether this solution satisfies the inequality constraints, which require that $\theta_c^* \geq 0$ for all c. Since $N \geq 0$ and $\alpha > 0$, the denominator of each θ_c^* is strictly positive. Similarly, since $\sum_{n=1}^{N}[x_n = c] \geq 0$ for each

c and $\alpha > 0$, the numerator is also strictly positive. This means that $\theta_c^* > 0$ for all c, which satisfies the inequality constraint, indicating that the obtains θ_c^*'s are the constrained global maximum of the penalized objective function.

Rubric: We were expecting a solution using the method of Lagrange multipliers. Solutions of this type should be graded as follows:

- 10 points: correctly obtained gradient of Lagrangian with respect to θ_c
- 10 points: correctly obtained gradient of Lagrangian with respect to normalization Lagrange multiplier.
- 10 points: correctly solved for θ_c in terms of data, α and normalization Lagrange multiplier.
- 10 points: correctly solved for value of normalization Lagrange multiplier in terms of data, C, and α.
- 10 points: correctly solved for θ_c in terms of data and α and C.
- -5 points: minor sign or other errors in derivation.

Correct solutions using the Lagrangian with correct work, but in a different way, are also worth full points. Correct solutions with correct work that do not use the Lagrangian are worth a maximum of 70 points.