Introduction to Machine Learning: Module 2

Ehsan Shaghaei

Ehsan Shaghaei, Applied Artificial Intelligence Dept., Innopolis University
e.shaghaei@innopolis.university

1 Introduction

In this document, two theoretical questions on K-means clustering and SVM were solved, besides the solutions for on classification and clustering were evaluated and purposed using CNN methods.Also the optional task on GANs were solved and the solution were provided in details.

2 Theoretical Tasks

2.1 Theoretical question on K-means Clustering (10%)

Firstly, since this part is about the globally optimal solution we know the case $D_1 = \{-2, -2, -2, ..., -2, a\}, D_2 = \{0, 0, 0, ..., 0\}$ is not feasible cause it results t0 a cluster with higher variance since $a > 0$ and $\|a - (-2)\| > \|a - 0\|$, thus we may skip checking this case as we know it will result to a bigger cost in comparison to the case below.
Nevertheless, for the case [1]:

$$\begin{cases} D = \{-2, -2, -2, ..., -2, 0, 0, 0, ..., 0, a\};\ n(D) = 2m + 1 \\ D_1 = \{-2, -2, -2..., -2\};\ n(D_1) = m \\ D_2 = \{0, 0, 0..., 0, a\};\ n(D_2) = m + 1 \\ D_1 \cup D_2 = D \\ D_1 \cap D_2 = \phi \end{cases}$$

$$\begin{cases} \mu_1 = \frac{\sum_{d \in D_1} d}{n(D_1)} = \frac{-2 \times m}{m} = -2 \\ \mu_2 = \frac{\sum_{d \in D_2} d}{n(D_2)} = \frac{0 \times m + a}{m+1} = \frac{a}{m+1} \end{cases}$$

$$J(D, D_1, D_2, \mu_1, \mu_2) = \sum_{i=1}^{2} \sum_{x \in D_i} \|x - \mu_i\|^2$$

[1] $n(S)$ represents the number of element in multi-set S

$$= \sum_{x \in D_1} \|x - \mu_1\|^2 + \sum_{x \in D_2} \|x - \mu_2\|^2$$

$$= (m \times ((-2) - (-2))^2) + (m \times (0 - \frac{a}{m+1})^2 + (a - \frac{a}{m+1})^2)$$

$$= 0 + m \times \frac{a^2}{(m+1)^2} + \frac{(am)^2}{(m+1)^2}$$

$$= \frac{a^2m^2 + a^2m}{(m+1)^2} = \frac{(a^2m)(m+1)}{(m+1)^2} = \frac{a^2m}{m+1}$$

$$MinJ; constraint\ a^2 < f(m)$$

$$\xrightarrow{Lagrange\ multiplier[1]} \psi(m, a, \lambda) = \frac{a^2m}{m+1} - \lambda(a^2 - f(m))$$

$$\begin{cases} \frac{\partial \psi}{\partial \lambda} = a^2 - f(m) = 0 \rightarrow a^2 = f(m) \\ \frac{\partial \psi}{\partial a} = \frac{2am}{m+1} - 2\lambda a = 0 \rightarrow \begin{cases} m = \frac{\lambda}{1-\lambda} \\ \lambda = \frac{m}{m+1} \end{cases} \\ \frac{\partial \psi}{\partial m} = \frac{a^2}{(m+1)^2} + \lambda \frac{\partial f(m)}{\partial m} = 0 \end{cases}$$

$$\xrightarrow[\lambda = \frac{m}{m+1}]{a^2 = f(m)} \frac{1}{(m+1)^2} \times f(m) + \frac{m}{m+1} \times \frac{\partial f(m)}{\partial m} = 0 \qquad (1)$$

$$\xrightarrow{\text{First Order linear ODE}} \frac{df(m)}{dm} = -\frac{1}{m(m+1)} f(m) \rightarrow \int \frac{\frac{df(m)}{dm}}{f(m)}.dm = \int -\frac{1}{m^2+m}.dm$$

$$\rightarrow f(m) = e^{C_1} \frac{m+1}{m}$$

2.2 Theoretical question on SVM (20%)

1. $\min_\theta \frac{1}{2}\|\theta\|^2$ subject to $y_t \theta^T x_t \geq 1$ for all $t \in \{1, ..., n\}$.[2]

   (a) **Can** be produced by this classifier.
   (b) **Can not** be produced by this classifier because in this case the hyper-plane does not go through origin while $\theta_0 = 0$.
   (c) **Can not** be produced by this classifier because in this case the constrain were not relaxed and points are not supposed to lie on the wrong side of the margin (miss-classification).
2. $\min_\theta \frac{1}{2}\|\theta\|^2$ subject to $y_t \theta^T (x_t + \theta_0) \geq 1$ for all $t \in \{1, ..., n\}$.

   (a) **Can not** be produced by this classifier since θ_0 should become zero but this won't be the solution for this case regarding the optimization problem.
   (b) **Can** be produced by this classifier, and it is the case of hyper-plane with the maximum margin.

(c) **Can not** be produced by this classifier because in this case the constrain were not relaxed and points are not supposed to lie on the wrong side of the margin (miss-classification).

3. For fixed $C \in (0, \infty), \min_{\theta,\zeta} \frac{1}{2}\|\theta\|^2 + C\sum_{t=1}^{n} \zeta_t$ subject to $y_t\theta^T x_t \geq 1 - \zeta_t$ and $\zeta_t \geq 0$ for all $t \in \{1, ..., n\}$.

   (a) **Can** be produced by this classifier in case a large value for regularization parameter C were chosen.
   (b) **Can not** be produced by this classifier, since θ_0 is not available for this case to be produced
   (c) **Can** be produced by this classifier in case a small value for regularization parameter C were chosen.

References

1. Lagrange multiplier. https://doi.org/10.1093/oi/authority.20110803100047903, https://www.oxfordreference.com/view/10.1093/oi/authority.20110803100047903
2. Cortes, C., Vapnik, V.: Support-vector networks. Machine learning **20**(3), 273–297 (1995)