Pattern Recognition and Machine Learning, January 22, 2020

Name:

1. (5 points) Describe the K-means algorithm (1pt). Prove that K-means converges to a local minimum (2pts). Describe the K-means++ algorithm (1pt). Give an example when employing K-means++ leads, on average, to better solution than K-means (1pt).

2. (5 points) Formulate the logistic regression problem (1 point). Derive the gradient descent learning method for this task (2.5 points).
   Given the training set $T = \{(-1,-1), (-2,-1), (0,1), (2,1)\}$, starting point $w = (0,0)$, and an initial step size of 1, do the first iteration of gradient descent. (1.5 points).

3. (5 points) Parameter Estimation. The reliability of light bulbs is defined by the probability density function

   $$p(t) = \frac{1}{\theta\left(\frac{t}{\theta}+1\right)^2}, \quad t \in [0, \infty],$$

   where $p(t)\delta$ is the probability that a bulb fails in a short time interval $(t, t+\delta)$. That is, the probability $P(T)$ that the bulb fails in interval $(0,T)$, is
   $P(T) = \int_0^T p(t)\mathrm{d}t = 1 - \frac{1}{(\frac{T}{\theta}+1)}$. What is the maximum likelihood estimate of the reliability parameter θ if in an experiment with two bulbs the following lifetimes have been observed:
   $t_1 = 1.0, t_2 = 4.0$? (4pts)
   For this setting, what is the probability that a light bulb is still OK at time 10? (1pt)

4. (5 points) Decision tree classifiers.
   (a) (2 points) Describe the simplest top-down tree-building algorithm that selects the decision rule at a node by maximization of information gain.
   (b) (3 points) Discuss advantages and disadvantages of decision trees.

(3) $L(\theta) = \prod_{t \in T} p(t) = \frac{1}{\theta\left(\frac{t_1}{\theta}+1\right)^2} \cdot \frac{1}{\theta\left(\frac{t_2}{\theta}+1\right)^2}$

$$\ell(\theta) = \log\left(\frac{1}{\theta\left(\frac{t_1}{\theta}+1\right)^2}\right) + \log\left(\frac{1}{\theta\left(\frac{t_2}{\theta}+1\right)^2}\right)$$

$$\ell(\theta) = \overset{0}{\cancel{\log(1)}} - \log\left(\theta\left(\frac{t_1}{\theta}+1\right)^2\right) + \overset{0}{\cancel{\log(1)}} - \log\left(\theta\left(\frac{t_2}{\theta}+1\right)^2\right)$$

$$\ell(\theta) = -\log\theta - 2\log\left(\frac{t_1}{\theta}+1\right) - \log\theta - 2\log\left(\frac{t_2}{\theta}+1\right)$$

$$\ell(\theta) = -2\log\theta - 2\log\left(\frac{t_1+\theta}{\theta}\right) - 2\log\left(\frac{t_2+\theta}{\theta}\right)$$

$$\ell(\theta) = \cancel{-2\log\theta} - 2\log(t_1+\theta) \cancel{+2\log\theta} - 2\log(t_2+\theta) + 2\log\theta$$

$$\ell(\theta) = 2\log\theta - 2\log(t_1+\theta) - 2\log(t_2+\theta)$$

$$\frac{\partial \ell}{\partial \theta} = \frac{2}{\theta} - \frac{2}{t_1+\theta} - \frac{2}{t_2+\theta} = 0 \quad \Big| \cdot \theta \cdot (t_1+\theta)(t_2+\theta)$$

$$\frac{\partial \ell}{\partial \theta} = 2(t_1+\theta)(t_2+\theta) - 2\theta(t_2+\theta) - 2\theta(t_1+\theta) = 0$$

$$= 2(t_1t_2 + t_1\theta + t_2\theta + \theta^2) - 2\theta t_2 - 2\theta^2 - 2\theta t_1 - 2\theta^2$$

$$= 2t_1t_2 + \cancel{2t_1\theta} + \cancel{2t_2\theta} + \cancel{2\theta^2} \cancel{-2\theta t_2} \cancel{-2\theta^2} \cancel{-2\theta t_1} - 2\theta^2$$

$$2t_1t_2 - 2\theta^2 = 0$$

$$2\theta^2 = 2t_1t_2$$

$$\theta^2 = t_1t_2$$

$$\theta = \pm\sqrt{t_1t_2} = \pm\sqrt{4} = \underline{\underline{\pm 2}}$$

(h) $P(T) = 1 - \frac{1}{\left(\frac{T}{\theta}+1\right)} = \int_0^T p(t)\,dt \longrightarrow$ fails in $0 \ldots T$

$$\text{ok} @ T = 1 - \left(1 - \frac{1}{\frac{T}{\theta}+1}\right) = \frac{1}{\frac{T}{\theta}+1} = \frac{1}{\frac{10}{2}+1} = \underline{\underline{\frac{1}{6}}}$$

(2) $$\underbrace{a(x) = \ln \frac{p(1|x)}{p(2|x)}}_{\text{log odds}} = w^T \cdot x \to \begin{bmatrix} x_0 = 1 \\ x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$[\underbrace{w_0}_{\text{bias}} \ldots w_D]$$

$$p(1|x) = \frac{1}{1+e^{-wx}} \quad h = 1$$

$$p(k|x) = \frac{1}{1+e^{-hwx}}$$

$$p(2|x) = \frac{1}{1+e^{wx}} \quad h = -1$$

$$l(w) = \sum_{i=1}^{N} \log p_w(x_i, h_i) = \sum_{i=1}^{N} \log p(h_i|x_i) + \cancel{\log(p(x_i))}$$

nezávisí na minimalizaci

$$= - \sum_i \log\left(1 + e^{-h_i w x_i}\right) \quad \text{minimalizace}$$

$$q(x) = \begin{cases} 1 & \text{if } w \cdot x > 0 \\ 2 & \text{else} \end{cases}$$

$w = (0,0)$

$$\frac{\partial l}{\partial w} = \sum_i \frac{e^{-h_i w x_i} \cdot h_i x_i}{1 + e^{-h_i w x_i}} \qquad \frac{h_i x_i}{2}$$

$$T = \{(\overset{x}{-1}, \overset{h}{-1}), (-2,-1), (0,1), (2,1)\} \qquad x_i = \begin{pmatrix} 1 \\ x_i \end{pmatrix}$$