CS589: Machine Learning - Fall 2017
Quiz 5

Dec. 4, 2017 Name: Jie Wang

Instructions: You have 30 minutes to complete this quiz. No calculator is permitted. You may answer any requiring a numerical response with an unevaluted expression, i.e write "$\frac{2\times 10}{100}$" rather than ".2". This expression can contain only *specific numbers*, e.g. you may not write "2^2" but not "$2K$" since that contains "K". Only the final form will be graded. If you perform intermediate calculations, draw a box around your final answer. Each minute late the quiz is turned in will result in a 20 point penalty.

1. Bayesian Probabilities (30 points)

Recall the setup of the "Monkeys" problem from class:

$$
\begin{aligned}
Pr(M = a) &= .25 \\
Pr(M = b) &= .75 \\
Pr(B = g|M = a, W = c) &= .8 \\
Pr(B = g|M = a, W = r) &= .2 \\
Pr(B = g|M = b, W = c) &= .2 \\
Pr(B = g|M = b, W = r) &= .8
\end{aligned}
$$

Question: Suppose you observe two rainy days, and two green balls, i.e. $W = rr$ and $B = gg$. What is the probability that Alfred is the monkey, i.e. what is $Pr(M = a|W = rr, B = gg)$?

$$P(M=a|W=rr, B=gg) = \frac{P(M=a)P(W=rr, B=gg|M=a)}{P(W=rr, B=gg)} = \frac{P(W=rr)\, P(M=a)P(B=gg|M=a, W=rr)}{P(W=rr, B=gg)}$$

$$\boxed{\frac{1}{49}}$$

3. K-means (30 points)

Recall the K-means algorithm. After initialization, one iterates between the two steps of

- Fix $m_1, m_2, ..., m_K$ and set $C(i) = \text{argmin}_{1\leq k\leq K} \|x_i - m_k\|_2^2$ for all i.
- Fix C and set $m_k = \frac{1}{N_k}\sum_{i:C(i)=k} x_i$ for all k,

where $N_k = \sum_{i=1}^{N} I[C(i) = k]$ is the number of points currently assigned to cluster K. Suppose that we have the following one-dimensional dataset and the current state for C:

i	1	2	3	4	5	6
x_i	0.5	2.0	3.0	5.0	5.5	7.0
$C(i)$	3	1	2	2	3	2

Question: If we were to run the step to compute the new values of m_k, what would they be?

$$
\begin{aligned}
m_1 &= 2 \\
m_2 &= 5 \\
m_3 &= 3
\end{aligned}
$$

3. A change to the Metropolis Algorithm (40 points)

Recall the metropolis algorithm. One takes as input an unnormalized distribution $\hat{p}(m)$ and a symmetric proposal distribution q and uses the following steps

- Initialize m^0
- For $t = 0, 1, ..., T-1$
  - $m' \sim q(m'|m^t)$
  - If rand() $\leq \hat{p}(m')/\hat{p}(m^t)$
    - $m^{t+1} = m'$
  - Else
    - $m^{t+1} = m^t$
- Return $m^1, m^2, ..., m^T$.

We saw in class that the probability of transitioning from state m to state n was

$$Pr(m \to n) = q(n|m) \min(1, \hat{p}(n)/\hat{p}(m)).$$

We also saw that these transitions satisfy detailed balance with respect to p, i.e. that for each pair of states m and n,

$$p(m)Pr(m \to n) = p(n)Pr(n \to m).$$

New algoroithm: Now, consider an alternative version of the algorithm which is identical in all respects, except that the fourth line of the algorithm is changed:

- Initialize m^0
- For $t = 0, 1, ..., T-1$
  - $m' \sim q(m'|m^t)$
  - If rand() $\leq \frac{\hat{p}(m')}{\hat{p}(m')+\hat{p}(m^t)}$
    - $m^{t+1} = m'$
  - Else
    - $m^{t+1} = m^t$
- Return $m^1, m^2, ..., m^T$.

Question: In the alternate algorithm, what is the probability of transitioning from state m to state n?

$$Pr(m \to n) = q(n|m) \min\left(1, \frac{\hat{p}(n)}{\hat{p}(n)+\hat{p}(m)}\right)$$

Question: Does this alternative algorithm also satisfy detailed balance with respect to p? Answer clearly "yes" or "no" and give a short justification of your answer.

Yes

Since $$\frac{Pr(m\to n)}{Pr(n \to m)} = \frac{q(n|m)\min\left(1, \frac{\hat{p}(n)}{\hat{p}(n)+\hat{p}(m)}\right)}{q(m|n)\min\left(1, \frac{\hat{p}(m)}{\hat{p}(m)+\hat{p}(n)}\right)} = \frac{\hat{p}(n)/(\hat{p}(n)+\hat{p}(m))}{\hat{p}(m)/(\hat{p}(m)+\hat{p}(n))} = \frac{\hat{p}(n)}{\hat{p}(m)},$$

$$\Rightarrow Pr(m\to n)\cdot \hat{p}(m) = Pr(n \to m)\hat{p}(n).$$