Data Mining Quiz March 2, 2016

1. Recall that *gradient descent* is an iterative optimization method for determining local minima. The update rule for gradient descent is $x_{i+1} = x_i - \gamma \nabla f(x_i)$. Suppose you have the objective function $f(x) = x^2 - 6x + 10$ to which you apply this algorithm for an initial point $x_0 = 5$. For which values of γ is it true that $f(x_1) \leq f(x_0)$ (assume $\gamma \geq 0$)?

2. What is the maximum possible *rank* of a 3×2 matrix?

3. The singular value decomposition and non-negative matrix factorization are algorithms for computing a factorization of a matrix, that is writing a given matrix as a product of other matrices (factors). Describe the properties of the factors in each case, SVD and NMF.

4. What are the *singular values* of a matrix?

   Two matrices $\mathbf{A}$ and $\mathbf{B}$ are shown below. The first singular value of $\mathbf{A}$ is $\sigma_0 \approx 5.29$. What is the next singular value (provide your reasoning)? Answer the same question for $\mathbf{B}$ and $\sigma_0 \approx 3.873$. In either case, if you cannot answer the question exactly an approximate answer is acceptable.

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 2 & 2 \\ 3 & 3 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 2.001 \end{bmatrix} \tag{1}$$

5. What is the *document-term* matrix?

6. Would you classify Latent Semantic Analysis as a supervised or unsupervised learning technique? Explain your reasoning.

7. Suppose you have a "document corpus" consisting of the following three sentences:

   - (a) *The dog barked.*
   - (b) *I like chewing gum.*
   - (c) *The cat meowed.*

   Find the document-term matrix for this corpus.

8. Based on your answer from the previous problem, use cosine similarity to make document-document or term-term comparisons. Which pair of *documents* is most similar? Which pair of *terms* is most similar?

9. What is the free parameter in non-negative matrix factorization and what is the objective function?

10. A quadratic form is an expression like $\mathbf{x}^T\mathbf{A}\mathbf{x}$ for vector $\mathbf{x}$ and matrix $\mathbf{A}$. Assume quadratic forms are analogous to quadratic functions from single variable calculus. What do expect is the $\mathbf{x}$ dependence of $\frac{\partial^2}{\partial \mathbf{x}^2}\mathbf{x}^T\mathbf{A}\mathbf{x}$?