Re-Test

- Do not refer to Internet resources. Do not discuss with class mates or your friends during the test/quiz.
- Write on one or two A4 pages. Write brief; Precise. Scan and send back within the time.
- Make assumptions if really required. But state.

1. Write the expression for sigmoid and show that its derivative can be expressed in terms of sigmoid itself.
2. Why is vanishing/exploding gradient problem very serious in recurrent neural networks? (provide brief and precise answer)
3. We know LDA for two class problem. Motivated by this, how do we extend the objective to three class. Write the expression (i.e., objective function) and define your notation. (no need to derive solution!).
4. Consider a kernel $\kappa(x, y) = \phi(x)^T \phi(y)$. Show that Kernel matrix (i.e., K) is PSD.
5. It is argued that an MLP with one hidden layer, but with no activation/nonlinearity is only as good as a SLP. If you agree, show analytically with the help of an MLP of 3 input, one hidden of 8 neurons and one output of 2 neurons.
6. Argue that the LOO error estimate of SVM is $\frac{|SV|}{N}$ where N is the training data size and SV is the set of support vectors.
7. Consider K-Means with $K = 2$. To initialize we start with $C_1 = \{[1, 1]^T, [-1, -1]^T, [2, 2]^T, [-2, -2]^T, [3, 3]^T\}$ and $C_2 = \{[-1, 1]^T, [1, -1]^T, [-2, 2]^T, [2, -2]^T\}$. Run two iterations of K-Means. Show means and partitions.
8. Different types of convolutions are popular. We also see 1×1 convolutions in todays deep neural networks. Is this an Identity mapping? Why is this useful?
9. We know that K-Means minimizes $\sum_{i=1}^{K} \sum_{x_j \in C_i} ||x_j - \mu_i||_2^2$. Assume a lucky implementation where $K-Means$ is initialized well and we reach global minima in every attempt.

   Let us start with $N = 1000$ samples we run $K - Means$ with K ranging from 2 to 100. Consider a measure $J = \frac{1}{K} \sum_{i=1}^{K} \sum_{x_j \in C_i} ||x_j - \mu_i||_2^2$. Plot K vs J conceptually. Explain in 2 sentence on the trend.
10. Consider an MLP trained for a classification problem. Hidden layers have sigmoid activation.

    We have enough samples from two classes. We use a cross entropy loss. Conceptually plot Loss vs iteration for (i) Training set (ii) Validation set (iii) Test Set. Explain in 2 sentence on the trend.