111學年第二學期第一次考試　科目：機器學習

班級：資管碩一　**應考人數**：28　**日期**：112/03/30　**考試時間**：2:10-4:00

注意：可使用計算器，不可使用任何字典。

1. What method does the backpropagation algorithm use to train a neural network? (5%)
(a) Maximum likelihood estimation (b) Error propagation method (c) Negative log likelihood estimation (d) Gradient-decent method.

2. If **x** = [-2 4] and **y** = [1 3], what is **L$_1$ norm** of (**x** - **y**)? (5%)
(a) 2 (b) 5 (c) 4 (d) 10

3. Which optimizer **does not** individually adapt the learning rates of all model parameters by scaling them inversely proportional to the square root of the sum of all of their historical squared values? (5%)
(a) SDG with momentum (b) AdaGrad (c) RMSProp (d) Adam.

4. Which statements are correct? (5%)
(a) Shallow neural networks may overfit more.
(b) Regularization techniques can be used to reduce model overfitting.
(c) One hidden layer is enough to learn an approximation of any function to an arbitrary degree of accuracy
(d) A network with a deeper depth seems to result in better generalization for a wide variety of tasks

5. What steps are included in a machine learning pipeline? (5%)
(a) Report writing (b) Feature generation (c) Data cleaning (d) Model selection.

6. Please explain the following terms. (Note: do not just translate) (25%)
(a) early stopping with patience = 10 (b) batch size = 64 (c) epoch = 20 (d) bagging (e) multi-task learning.

7. Assume batch size is 3 and the dimension of each data point is 4. Please use an example to explain and show the difference between **batch normalization** and **layer normalization**. (10%)

8. Assume the output of a hidden layer before the activation function is [3 2 1 3 6 1] and the activation is **softmax**. What is the **output vector** of the softmax activation function? (10%)

$$\text{softmax}(\mathbf{x}) = \frac{1}{\sum_{j=1}^{K}\exp(x_j)}\begin{bmatrix}\exp(x_1)\\ \exp(x_2)\\ \dots\\ \exp(x_k)\end{bmatrix} \quad \text{where} \exp(x_i) = e^{x_i} \text{ and } e = 2.718$$

9. Assume vector **a** = (1, 3, 1) and vector **b** = (2, 2, 1). What is the **scalar project** of **a** onto **b** and the **vector projection** of **a** onto **b**? (10%)
Hint: scalar project of a onto b: $\mathbf{a}\cdot\frac{\mathbf{b}}{||\mathbf{b}||}$, vector project of **a** onto **b**: $\left(\mathbf{a}\cdot\frac{\mathbf{b}}{||\mathbf{b}||}\right)\frac{\mathbf{b}}{||\mathbf{b}||}$

10. Assume batch size is 3, the output of the batch from a neural network is (0.8, 0.2, 0.7), and the ground truths is (1, 0, 0). What is the **binary cross entropy**? (10%)
Hint: Binary cross entropy: loss = $-\frac{1}{N}\sum_{n=1}^{N} y_n\log\hat{y}_n + (1-y_n)\log(1-\hat{y}_n)$

11. If **x** = [-2 4] and **y** = [1 3], what is **L$_2$** norm and **max norm** of (**x** – **y**)? Namely, $||\mathbf{x}-\mathbf{y}||_2$ and $||\mathbf{x}-\mathbf{y}||_\infty$ (10%)
Hint: $||\mathbf{x}||_p = (\sum_i |x_i|^p)^{\frac{1}{p}}$, and $||\mathbf{x}||_\infty = \max_i |x_i|$