Week 5 Practice

CSCI 567 Machine Learning

Spring 2025

Instructor: Haipeng Luo

1. MULTIPLE-CHOICE QUESTIONS: One or more correct choice(s) for each question.

1.1. Which of the following is/are **true** about neural nets?
(A) A neural net with one hidden layer and a fixed number of neurons can represent any continuous function.
(B) A fully connected feedforward neural net without nonlinear activation functions is the same as a linear model.
(C) Dropping random neurons in each iteration of Backpropagation helps prevent overfitting.
(D) A max-pooling layer with a 2×2 filter has 4 parameters to be learned.

1.2. Which of the following can help prevent overfitting in neural nets?
(A) Retraining on the same data many times.
(B) Using a validation set for early stopping.
(C) Data augmentation.
(D) Training until you get the smallest training error.

1.3. Suppose a convolution layer takes a 8×8 image with 3 channels as input and outputs a $4 \times 4 \times 8$ volume. Which of the following is a possible configuration of this layer?
(A) One 4×4 filter with depth 8, stride 2, 1 pixel of zero-padding.
(B) Three 4×4 filters with depth 8, stride 2, no zero-padding.
(C) Eight 4×4 filters with depth 3, stride 2, 1 pixel of zero-padding.
(D) Eight 4×4 filters with depth 3, stride 1, 1 pixel of zero-padding.

1.4. How many parameters do we need to learn for the following network structure? An $32 \times 32 \times 3$ image input, followed by a convolution layer with 3 filters of size 3×3 (stride 1, 1 pixel of zero-padding), then another convolution layer with 4 filters of size 2×2 (stride 2, no zero-padding), and finally a max-pooling layer with a 2×2 filter (stride 1, no zero-padding). (Note: the depth of all filters are not explicitly spelled out, and we assume no bias/intercept terms are used.)

(A) 43
(B) 97
(C) 129
(D) 145

1.5. What is the final output dimension of the last question?

(A) $15 \times 15 \times 1$
(B) $16 \times 16 \times 4$
(C) $32 \times 32 \times 1$
(D) $15 \times 15 \times 4$

1.6. Which of the following is/are **true** about kernel functions?

(A) If k_1 and k_2 are valid kernels, then so is $c_1 k_1 + c_2 k_2$ for any $c_1, c_2 \geq 0$.
(B) Kernel functions must be symmetric, i.e., $k(\boldsymbol{x}, \boldsymbol{x}') = k(\boldsymbol{x}', \boldsymbol{x})$.
(C) If k is a kernel, then $-k$ is a kernel too.
(D) If k is a kernel, then $\ln(k)$ is a kernel too.

1.7. Which of the following are not kernel functions?

(A) $k(x, x') = (\boldsymbol{x}^T \boldsymbol{x}')^2$
(B) $k(x, x') = -\|\boldsymbol{x} - \boldsymbol{x}'\|_2^2$
(C) $k(x, x') = (\boldsymbol{x}^T \boldsymbol{x}' + 1)^2$
(D) $k(x, x') = \|\boldsymbol{x} - \boldsymbol{x}'\|_2^2$

1.8. Vovk's real polynomial kernel $k : \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}$ is defined as: $k(\boldsymbol{x}, \boldsymbol{x}') = \frac{1-(\boldsymbol{x}^T \boldsymbol{x}')^p}{1-(\boldsymbol{x}^T \boldsymbol{x}')}$, where p is a non-negative integer. Which of the following is the corresponding feature mapping for this kernel when $D = 2$ and $p = 2$?

(A) $\phi(\boldsymbol{x}) = [x_1, x_2, x_1 x_2]^T$
(B) $\phi(\boldsymbol{x}) = [x_1, x_2, x_1^2, x_2^2]^T$
(C) $\phi(\boldsymbol{x}) = [x_1, x_2, 1]^T$
(D) $\phi(\boldsymbol{x}) = [x_1, x_2, x_1^2, x_2^2, x_1 x_2]^T$