Machine Learning 2016
Final Exam

14 December 2016

Name: ....

Question 1: $\boxed{\text{2 points}}$ Consider a binary classification task in two dimensions (x_1, x_2), and let $\mathcal{D} = \{((0,1), +1), ((-2,1), +1), ((-2,-1), -1)\}$ be a dataset of input-output pairs. Let $\mathbf{w}_0 = [0, 1, 1]$ be the initial weight vector. Describe the first two updates performed by the Perceptron Learning Algorithm. What is the total classification error after these two updates?

Question 2: $\boxed{\text{1 point}}$ Two common approaches for supervised learning are linear regression and quadratic regression. Name two reasons why you might prefer linear regression, and two reasons why you might prefer quadratic regression.

Name: ....

Question 3: $\boxed{\text{1 point}}$ Name three techniques that are commonly used in practice to improve the out-of-sample error (i.e. generalization properties) of a given supervised learning algorithm.

Question 4: $\boxed{\text{1 point}}$ What is the bias-variance tradeoff? Describe one way to decrease the bias and one way to decrease the variance.

Question 5: $\boxed{\text{1 point}}$ The backpropagation algorithm is widely used to train the weights of feed-forward neural networks. Name one similarity and one difference between backpropagation and logistic regression.

Question 6: $\boxed{\text{1 point}}$ Name three differences between convolutional neural networks and standard feed-forward neural networks.

Name: ....

Question 7: $\boxed{\text{2 points}}$ Consider the problem of online learning with a hypothesis set H: in every round $t = 1, 2, \ldots, n$, the learner observes an input $\mathbf{x}_t$, predicts the label $\widehat{y}_t$, observes the true label y_t, and then incurs the 0-1 loss $\ell(\mathbf{x}_t, y_t, \widehat{y}_t) = |y_t - \widehat{y}_t|$. With this notation, define the *regret against hypothesis* h and the *worst-case regret*! What is the interpretation of regret in the realizable case when there exists a hypothesis with zero cumulative loss?

Question 8: $\boxed{\text{1 point}}$ Consider the Exponential Weights Algorithm (`EWA`) for online learning with a finite hypothesis class H of size $|H| = K$. For a parameter $\eta > 0$, the regret of `EWA` satisfies $\mathbb{E}[R_n] \le \frac{\log K}{\eta} + \frac{\eta n}{8}$. Give a choice of $\eta > 0$ that makes this regret bound grow sublinearly (i.e., satisfying $\lim_{n\to\infty} \mathbb{E}[R_n]/n = 0$)! What is the choice of η that minimizes the bound?

Question 9: $\boxed{\text{1 point}}$ Consider the `Exp3` algorithm for the problem of K-armed bandits. For a parameter $\eta > 0$, the regret of `Exp3` satisfies $\mathbb{E}[R_n] \le \frac{\log K}{\eta} + \frac{\eta K n}{2}$. Explain the difference between this regret bound and the regret bound of `EWA` for the full-information online learning setting. What is the intuitive reason for this difference? In particular, why is the bandit problem harder than the full-information online learning problem?

Name: ....

Question 10: $\boxed{\text{2 point}}$

- Draw a Bayesian network of (at least) five nodes. Illustrate each of the three canonical cases in the D-separation algorithm using three different independence queries involving three sets of evidence variables. Note that you don't need to specify the conditional probability tables in this exercise.

- Transform the previous Bayesian network into a factor graph. Write down the corresponding belief propagation equations for obtaining marginal probabilities.

Question 11: $\boxed{\text{2 point}}$ Assume that a variable **x** comes from a probability distribution of the form

$$p(\mathbf{x}|\mathbf{w}) = \frac{1}{Z(\mathbf{w})} \exp\left(\sum_k w_k f_k(\mathbf{x})\right),$$

where the functions $f_k(\mathbf{x})$ are given, and the parameters $\mathbf{w} = \{w_k\}$ are not known. A dataset $\{\mathbf{x}^{(n)}\}$ of N points is supplied.

Show by differentiating the log likelihood that the maximum-likelihood parameters $\mathbf{w}_{ML}$ satisfy

$$\sum_{\mathbf{x}} P(\mathbf{x}|\mathbf{w}_{ML}) f_k(\mathbf{x}) = \frac{1}{N} \sum_n f_k(\mathbf{x}^{(n)}).$$

Name: ....

Follow the two following steps

- Write down the log-likelihood $\ln P(\{\mathbf{x}^{(n)}\}|\mathbf{w})$

- Differentiate the log of the normalizing constant $Z(\mathbf{w})$ w.r.t $\mathbf{w}$ and solve for $\mathbf{w}$. Remember that $\frac{\partial \ln(g(w))}{\partial w} = \frac{1}{g(w)} \frac{\partial g(w)}{\partial w}$.

The total score is: ______________ /15.