Q1. Machine Learning: Potpourri

(a) What is the minimum number of parameters needed to fully model a joint distribution $P(Y, F_1, F_2, ..., F_n)$ over label $Y$ and $n$ features $F_i$? Assume binary class where each feature can possibly take on $k$ distinct values.

(b) Under the Naive Bayes assumption, what is the minimum number of parameters needed to model a joint distribution $P(Y, F_1, F_2, ..., F_n)$ over label $Y$ and $n$ features $F_i$? Assume binary class where each feature can take on $k$ distinct values.

(c) You suspect that you are overfitting with your Naive Bayes with Laplace Smoothing. How would you adjust the strength $k$ in Laplace Smoothing?

- Increase $k$
- Decrease $k$

(d) While using Naive Bayes with Laplace Smoothing, increasing the strength $k$ in Laplace Smoothing can:

- Increase training error
- Decrease training error
- Increase validation error
- Decrease validation error

(e) It is possible for the perceptron algorithm to never terminate on a dataset that is linearly separable in its feature space.

- True
- False

(f) If the perceptron algorithm terminates, then it is guaranteed to find a max-margin separating decision boundary.

- True
- False

(g) In binary perceptron where the initial weight vector is $\vec{0}$, the final weight vector can be written as a linear combination of the training data feature vectors.

- True
- False

(h) For binary class classification, logistic regression produces a linear decision boundary.

- True
- False

(i) In the binary classification case, logistic regression is exactly equivalent to a single-layer neural network with a sigmoid activation and the cross-entropy loss function.

- True
- False

(j) You train a linear classifier on 1,000 training points and discover that the training accuracy is only 50%. Which of the following, if done in isolation, has a good chance of improving your training accuracy?

- Add novel features
- Train on more data

(k) You now try training a neural network but you find that the training accuracy is still very low. Which of the following, if done in isolation, has a good chance of improving your training accuracy?

- Add more hidden layers
- Add more units to the hidden layers
Q2. Neural Networks: Representation

For each of the piecewise-linear functions below, mark all networks from the list above that can represent the function \textbf{exactly} on the range $x \in (-\infty, \infty)$. In the networks above, $\text{relu}$ denotes the element-wise ReLU nonlinearity: $\text{relu}(z) = \max(0, z)$. The networks $G_i$ use 1-dimensional layers, while the networks $H_i$ have some 2-dimensional intermediate layers.