ECE542 - Quiz 3

1. [2 pts] In one sentence, what does the convergence theorem for the perceptron algorithm tell us about the number of steps needed for convergence? You don't have to write any formulas. Just intuition / interpretation is enough.
2. [1 pt] Which shape better describes a level set for the L_1 regularization penalty a circle, a diamond or a square?
3. [1 pt] When performing model averaging, what assumption about the correlation between the error of the models yields the smallest error variance?
4. [1 pt] Does dropout help backpropagation converge faster or does it help us have a more generalizable model?
5. [2 pt] When performing dropout, we virtually keep nodes in a layer with probability p. If after training we get a set of weights W, what values of weights should be used when testing a new sample?
6. [3 pts] Given that an adversarial sample for the differential strategy is generated by $x_a \rightarrow x + \epsilon \cdot sign\left(\nabla_x J(\theta, x, y)\right)$, where $J(\theta, x, y)$ is the standard cost used for training, then what is the expression for the modified cost $\tilde{J}(\theta, x, y)$ that incorporates the adversarial samples.

Answer

1. There is a finite number of updates needed to converge
2. Diamond
3. They are independent (or at least uncorrelated)
4. In general, more generalizable models
5. pW
6. $\tilde{J}(\theta, x, y) = \alpha \cdot J(\theta, x, y) + (1 - \alpha) \cdot J(\theta, x_a, y)$