1. Minimum Number of Binary classifiers required to classify N class classification problem
    a. log(N)
    b. ln(N)
    c. N
    d. ln(N)/ln(2)

    D Assign a number from 0-N-1 to a class and convert it to binary, Use one classifier to predict one bit, The setup is not optimal but one can solve the problem with this.

2. Does a perceptron always converge?
    a. Yes
    b. No
    c. No, in some cases.

    C converges for linearly separable data for others it need not converge.

3. You have observed oscillation in gradient descent of a convex function, What do you do to prevent/decrease oscillation?
    a. Increase learning rate
    b. Decrease learning rate
    c. Use second order methods to decide learning rate
    d. Give up hope

    B Decreasing learning rate would decrease step size and reduce oscillation.
    C Using Second order methods like Newton's method would scale the learning rate to prevent oscillation.

4. The output of a softmax activation function is bounded by
    a. 0,1
    b. -1,1
    c. It is unbounded
    d. None of the above

    A, B

5. Given a two class classification problem with 2 features per sample then.
    a. SVM linear == Naive Bayes Gaussian
    b. SVM linear > Naive Bayes gaussian
    c. SVM linear < Naive Bayes Gaussian
    d. Can't Say.

    C
    A linear svm can never learn a nonlinear decision boundary hence performs worse on non linearly separable data hence a Naive bayes with gaussian distribution performs better as it can learn
    When linearly separable data is present both perform equally well.