**CS412 – Machine Learning**

**Neural Networks**

**100pts**

1. **50pts -** We are trying to minimize a function F(**w**)= 5x2 –4x- 3xy + 10 with respect to the parameter **w**=[x y].

Starting from the point **w0**=[0 0], **use steepest descent algorithm** ***for TWO steps*** to find the local minimum of the function around this point. I.e. find **w1, w2** (corresponding to times t=1 and t=2) and corresponding F() values during the search. **You should use a “step size” alpha = 0.1.**

*Answer:*

*F(w0) =................................... : Just to note at what F value we start (5pts)*

*= ……………………………………………………………………: Compute the gradient (10pts)*

| ***w****0 = ………………………………..........: This is the gradient* ***evaluated*** *at w0 (10pts)*

***w****1=…………………………………………………………………………………….. :Update w0 to find w1 (5pts)*

*F(****w****1) =……………………………………………… : just checking to see if we are indeed minimizing*

***Now do the 2nd step as well: :20pts***

....

**2) 20pts –** Assume we have some children’s data points (Agei,Weighti) where **Weighti = 3.5\*Agei + 3 + ei** where ei are small, zero-mean noise (captures the deviations of the data from the above formula).

Show that a single neuron can learn this mapping by drawing the “network” including input, output, bias and activation function.

**3) 20pts –**

**a) 10pts** - When is a sigmoidal node (a neuron with sigmoid activation function) considered to be **saturated**? Check all that apply. No partial.

* when its net input is 0
* when its nets input is 10
* when its nets input is -10

**b) 10pts -** Explain your answer in a) **in one line** by referring to the derivative of its output w.r.t its net input.

**4) 10pts –** You have a 3-class classification problem (e.g. positive, negative, neutral), given a d-dimensional input. What is the number of output nodes and the activation function suitable for this?

.......................... nodes and ...................................................... activation function.