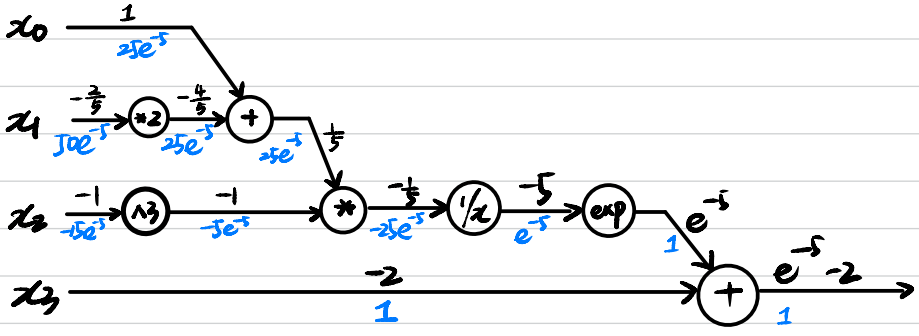
Name: YANG, tle Sid: 20901880 Hsc: hyangb2@connect.ust.hk

Q1:



$$\frac{\partial f}{\partial x_3} = 1$$

$$\frac{\partial f}{\partial x_0} = e^{\frac{1}{(x_0+2x_1)\cdot x_2^3}} \cdot \left(-\frac{1}{((x_0+2x_1)\,x_2^3)^2}\right) \cdot \left(x_2^3\right) = 25e^{-5}$$

$$\frac{\partial f}{\partial x_1} = e^{\frac{1}{(x_0+2x_1)\cdot x_2^3}} \cdot \left(-\frac{1}{((x_0+2x_1)\,x_2^3)^2}\right) \cdot \left(2x_2^3\right) = 50e^{-5}$$

$$\frac{\partial f}{\partial x_2} = e^{\frac{1}{(x_0+2x_1)\cdot x_2^3}} \cdot \left(-\frac{1}{((x_0+2x_1)\,x_2^3)^2}\right) \cdot (x_0+2x_1)\cdot 3x_2^2 = -15e^{-5}$$

Q2

1) Model with 2, 5, 10 hidden units.
   Underfit = fail to capture complex feature of data and thus model has low predicting accuracy.
2) Model with 50, 100 hidden units.
   Good = neither overfitting nor underfitting. Predicting accuracy remarkably increase to a satisfying result.
3) Model with 500, 1000 hidden units.
   Overfitting = match the dataset too precisely and may become inefficient for other dataset. Accuracy does not improve lot.