1. $P = p^5(1-p)^{25}$

$\ln(P) = \ln(p^5(1-p)^{25})$

$= 5\ln(p) + 25\ln(1-p)$

$f(x,y) = 5\ln(x) + 25\ln(y)$

$g(x,y) = x + y - 1 = 0$

$L(x,y,\lambda) = f(x,y) + \lambda g(x,y)$

$= 5\ln(x) + 25\ln(y) + \lambda(x+y-1)$

$\frac{\partial L}{\partial x} = \frac{5}{x} + \lambda = 0$

$\frac{\partial L}{\partial y} = \frac{25}{y} + \lambda = 0$

$\frac{\partial L}{\partial \lambda} = x + y - 1 = 0$

$\begin{cases} x = \frac{1}{6} \\ y = \frac{5}{6} \\ \lambda = -30 \end{cases}$

$\therefore \boxed{p = \frac{1}{6}}$

2. $f(x,y) = \sqrt{(x-0)^2 + (y-0)^2} = \sqrt{x^2+y^2}$ is equivalent to x^2+y^2

$g(x,y) = -\frac{1}{2}x - y + 5 = 0$

$L(x,y,\lambda) = f(x,y) + \lambda g(x,y)$

$= x^2 + y^2 + \lambda(-\frac{1}{2}x - y + 5)$

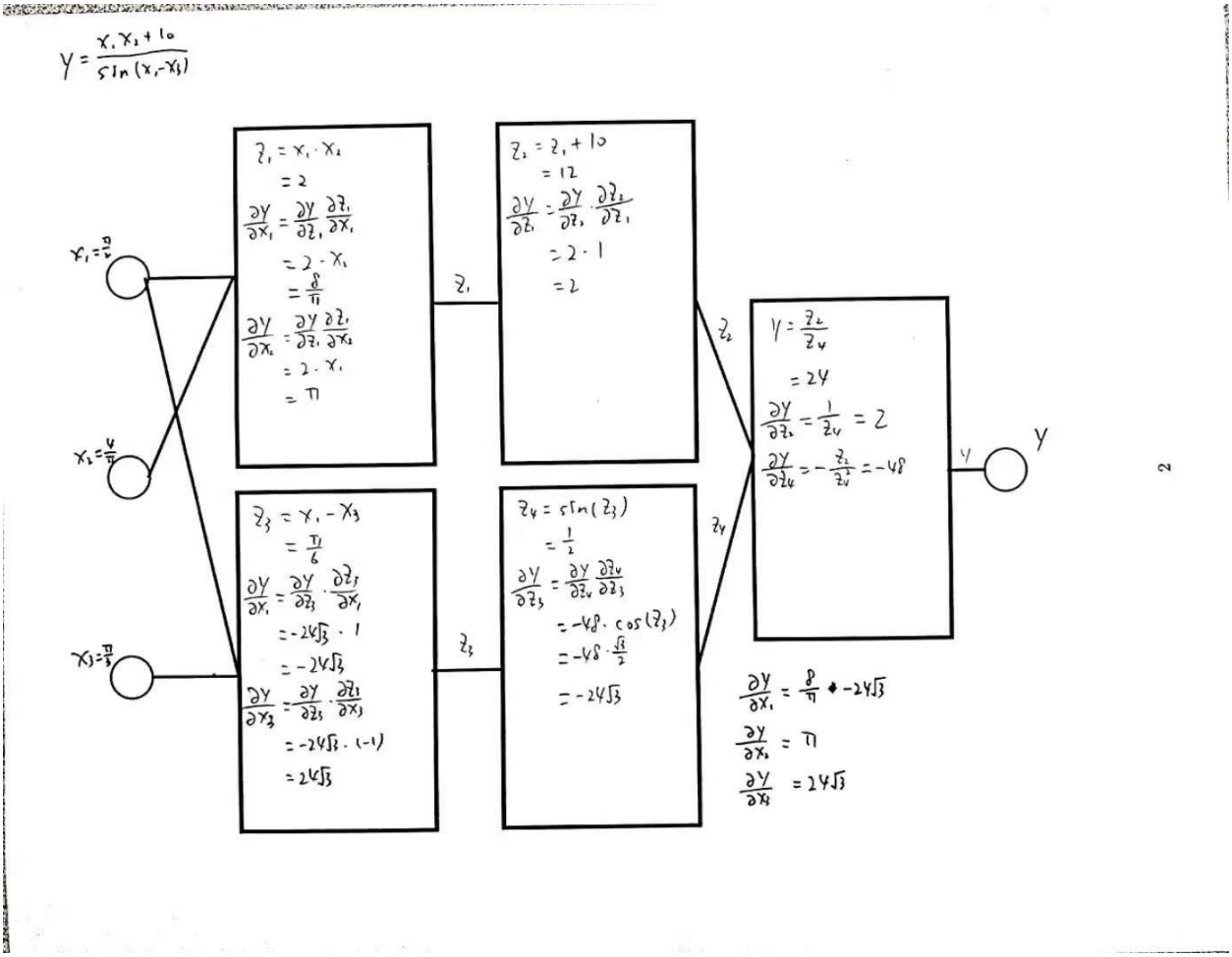
$\frac{\partial L}{\partial x} = 2x - \frac{1}{2}\lambda = 0$

$\frac{\partial L}{\partial y} = 2y - \lambda = 0$

$\frac{\partial L}{\partial \lambda} = -\frac{1}{2}x - y + 5 = 0$

$\begin{cases} x = 2 \\ y = 4 \\ \lambda = 8 \end{cases}$

$\therefore \boxed{\text{point } (2,4)}$



(a) What is currently driving deep learning?
Scale. Both on data and computation.

(b) What is end-to-end deep learning? Describe an end-to-end deep learning system for self driving cars? What is the limitation of end-to-end deep learning?
One end is the input and the other end is the output. It uses a Neural Network to go directly from the input to output. Its output has much more complex things than numbers. Usually it detects the positions of cars and pedestrians from the images. Then calculate the trajectory and get the steering direction. The use of end-to-end deep learning is to have the image and directly get the steering direction. But it needs really large amount of data. The data we have now is not enough.

(c) Classical machine learning is about the trade-off between bias (under-fitting) and variance (over-fitting). How is deep learning changing this trade-off?
The coupling between bias and variance is reduced. We have methods to improve one of them without reducing the other. Like a bigger model and have more data.

(d) Why is it a bad idea to use a validation set (dev set) that has a different distribution than the test set?

Validation set will be better if it matches the problem specification. If it is different than the testing set, the model will have many problems on the real testing because the validation set they used is under a different situation. It will lower the efficiency of working on the problem.

(e) Why is measuring human level performance for your problem useful?
Understand whether we should work on the bias part or the variance part. Human level performance acts as an approximation. It would be really hard to improve after surpass human level performance.

(f) Summarize Andrew Ng's personal advise on how to get started in machine learning research?
Go to Deep Learning School. Compete in Kaggle competitions. Read a lot of papers and make replicate results. Do dirty work. Keep doing it.