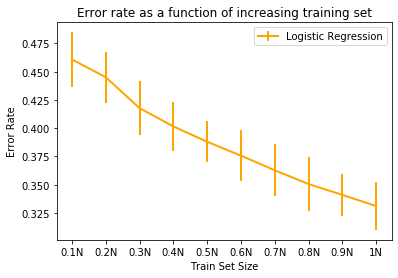
**Logistic Regression Stats:**



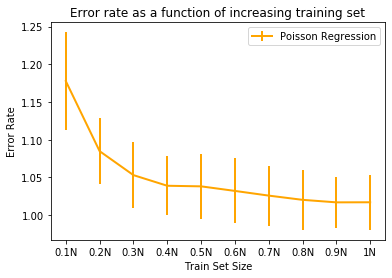
Average Convergence Time (in seconds):

[0.0013648351033528647, 0.002545785903930664, 0.005045318603515625, 0.0054384390513102215, 0.009569978713989258, 0.012146830558776855, 0.013626321156819662, 0.017661229769388834, 0.020917328198750813, 0.02614715099334717]

Average Number of Iterations:

[2.0, 2.0, 2.0, 2.0, 2.0, 2.0, 2.0, 2.0, 2.0, 2.0]

**Poisson Regression Stats:**



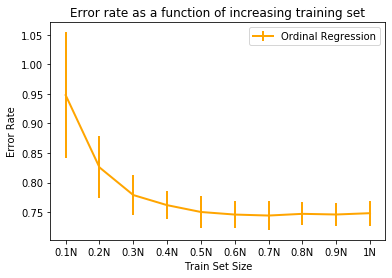
Average Convergence Time (in seconds):

[0.002909549077351888, 0.0031392256418863933, 0.010939860343933105, 0.01788318157196045, 0.021752556165059406, 0.028914229075113932, 0.03554205894470215, 0.043735655148824056, 0.05205761591593425, 0.06275574366251628]

Average Number of Iterations:

[6.166666666666667, 6.5, 6.6, 6.566666666666666, 6.6, 6.533333333333333, 6.466666666666667, 6.466666666666667, 6.366666666666666, 6.333333333333333]

**Ordinal Regression Stats:**



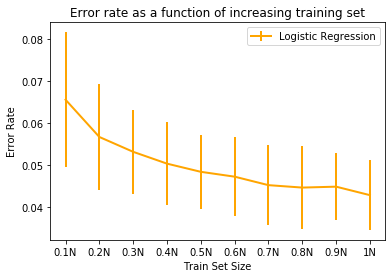
Average Convergence Time (in seconds):

[0.011883735656738281, 0.024133936564127604, 0.0382979154586792, 0.053017226854960124, 0.06425243218739828, 0.07947191397349039, 0.09363118807474773, 0.1094566027323405, 0.1263461669286092, 0.1371570348739624]

Average Number of Iterations:

[3.0, 3.0, 3.0, 3.0, 3.0, 3.0, 3.0, 3.0, 3.0, 3.0]

**Logistic Regression Stats: (For USPS data)**



Average Convergence Time (in seconds):

[0.01351478894551595, 0.017641393343607585, 0.02014126777648926, 0.027397871017456055, 0.040399996439615886, 0.05063343048095703, 0.06093369325002034, 0.07238805294036865, 0.08120932579040527, 0.08937756220499675]

Average Number of Iterations:

[4.2, 5.0, 5.0, 5.066666666666666, 5.5, 5.9, 6.0, 6.0, 6.0, 6.0]

**Observations:**

**Are the learning curves as expected?**

Yes, they are. Curves have decreasing error rate as our training size increases, which is expected. As the dataset size increases, error also has less deviation from mean error resulting in better performance.

**How does learning time vary across datasets for classification? and across the likelihood models?**

For Logistic, it takes only 2 iteration for wmap to converge and average 5 iterations for usps dataset.

For Poisson, it takes average 6.5 iterations and for Ordinal it takes average 3 iterations.

As the dataset size increases, the time it takes for wmap to converge increases. This is similar for all likelihood models.

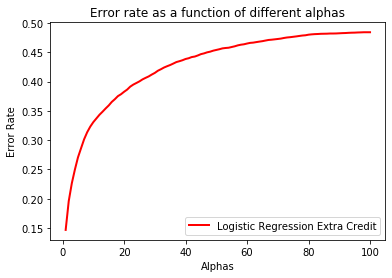
**What are the main costs affecting these (time per iteration, number of iterations)?**

Logistic performs the best out of all 3 since it has only two classes to predict. Also, Ordinal performs better than Poisson and converges in lesser iterations.

**Extra Credit**

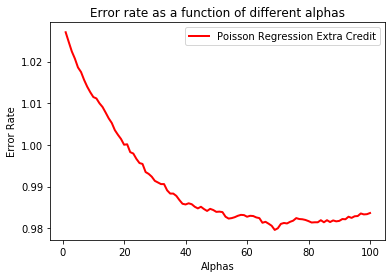
**Approach 1:**

To get the alpha, I tried iterating alpha value over range 1-100 and calculated mean error rate for each model. I also plotted error rate as a function of alpha values and time taken to compute this. **(Linear Search)**



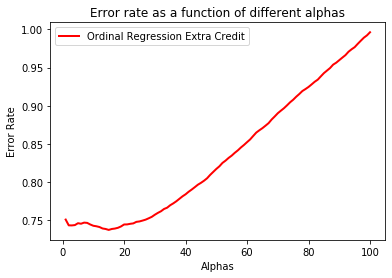
**Logistic**: Alpha where error rate is minimum: 1

Took 64.94578051567078 seconds



**Poisson**: Alpha where error rate is minimum: 69

Took 130.8319447040558 seconds

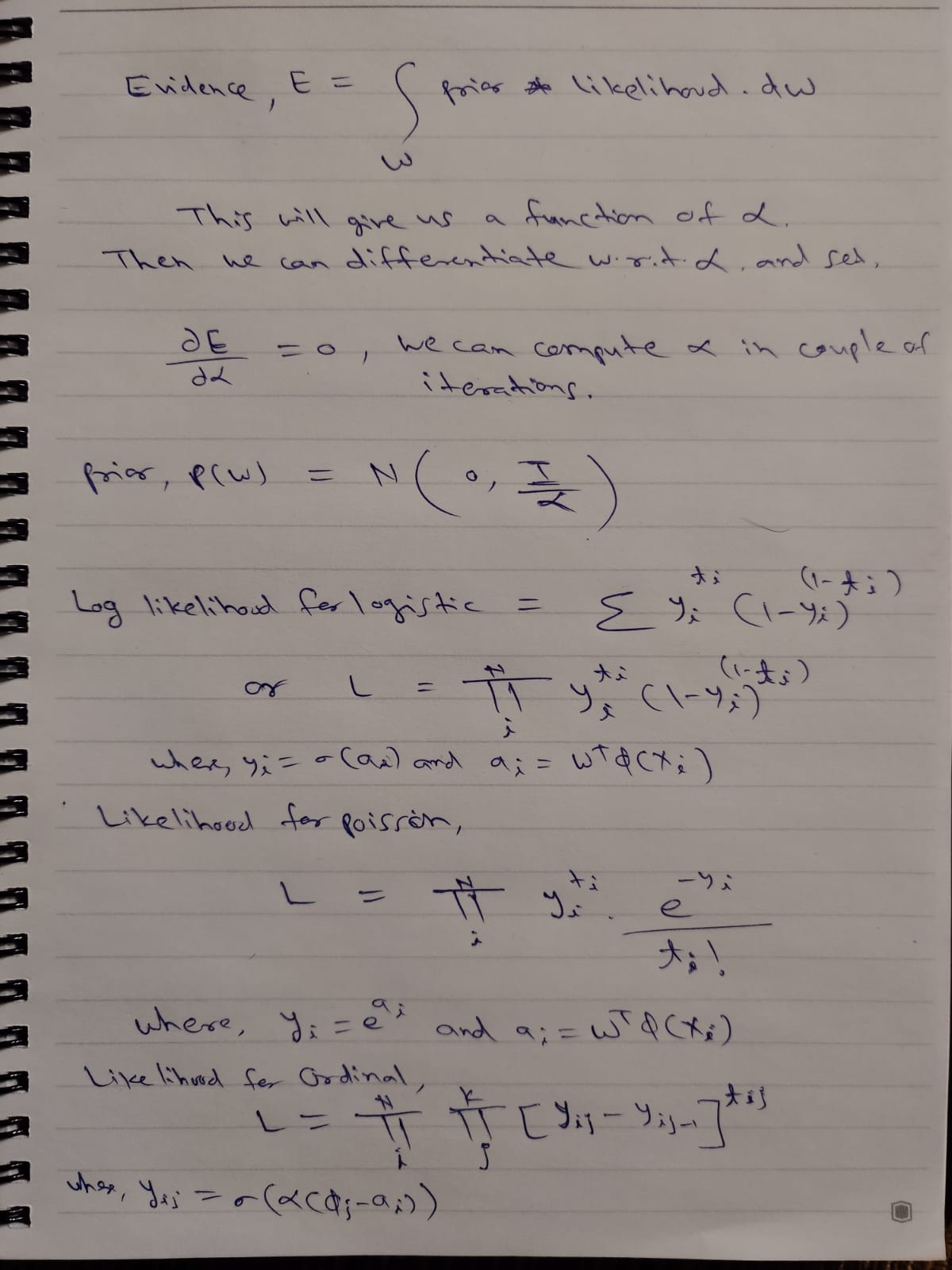


**Ordinal**: Alpha where error rate is minimum: 15

Took 459.042423248291 seconds

**Approach 2:**

We can use evidence function to calculate alpha. **Refer the image below:**

****

For each model, by calculating integral over prior \* likelihood, we get evidence equation which is in form of alpha. Then we can compute alpha in couple of iterations by differentiating using newtons method.

**I could not get time derive the equations for each model and implement it.**