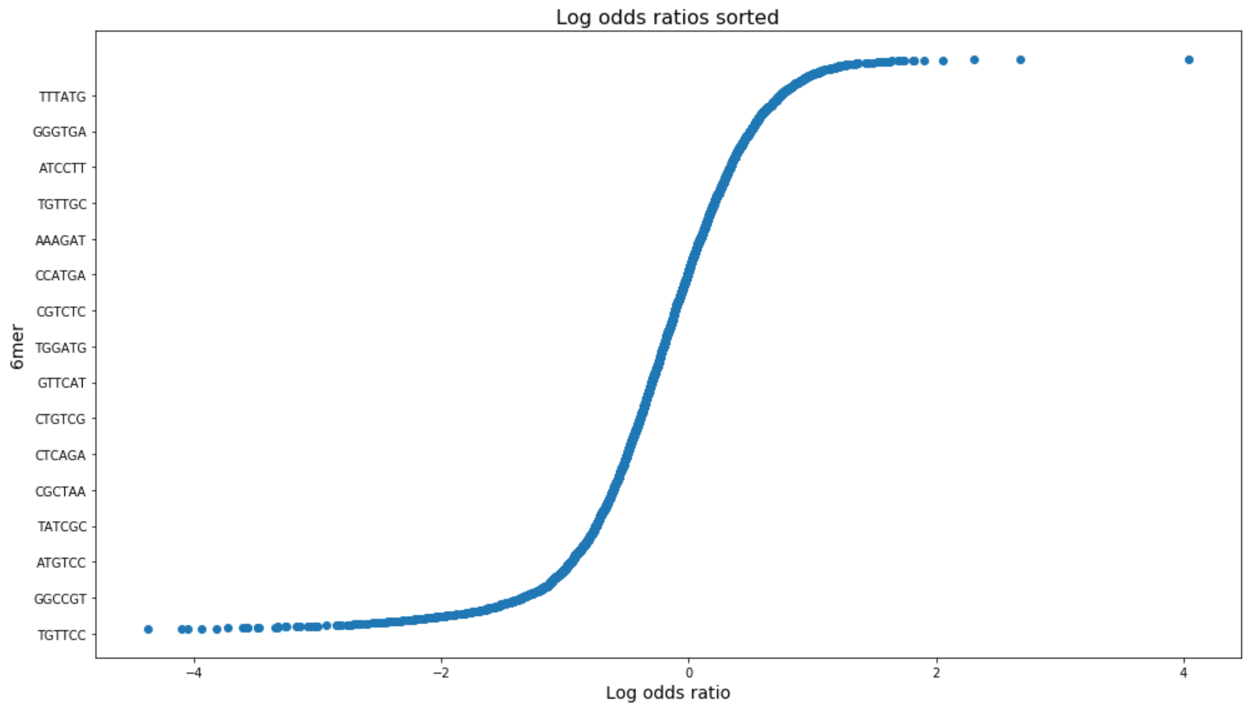
CSE520 Homework 4

Tony Adkins (anadkins@uw.edu)

Genomics

1. Sorted log odds ratios of 6mers



- Maximum log odds ratio: TCTTTC, 4.03856
- Minimum log odds ratio: CGACCT, -4.37356

The minimum above was calculated excluding the numerous 6mers which had zero probability, resulting in a negative infinity log odds ratio. The minimum of around -4 means the 6mer CGACCT has less than a 2:100 odds of splice usage. The maximum 6mer TCTTTC's log odds ratio of around 4 means something closer to 55:1, a nearly 3000-fold increase in odds likelihood.

2. KL-divergence Gradient Descent

∂L/∂wk derivation

$$\mathrm{L}(\vec{\mathrm{w}}, w_0, \lambda) = \frac{1}{\mathrm{N}} \sum_{i=1}^{N} \left(y_i * \log\left(\frac{y_i}{\hat{y}_i}\right) + (1 - y_i) * \log\left(\frac{1 - y_i}{1 - \hat{y}_i}\right) \right) + \frac{1}{2} \lambda \sum_{j=1}^{4096} {w_j}^2$$

Where y_i is a constant from the data and $\hat{y}_i$ is defined as

$$\hat{y}_i = \frac{1}{1 + e^{-(w_0 + \sum_{j=1}^{4096} w_j X_{i,j})}}$$

So, we can take the derivative as the sum of the derivatives of the two parts of the sum, starting with the first half:

$$\frac{\partial \frac{1}{\mathrm{N}} \sum_{i=1}^{N} \left(y_i * \log\left(\frac{y_i}{\hat{y}_i}\right) + (1 - y_i) * \log\left(\frac{1 - y_i}{1 - \hat{y}_i}\right)\right)}{\partial w_k}$$

Which can be further split into two, allowing us to find the derivatives of each separately because

$$\frac{1}{\mathrm{N}} \sum_{i=1}^{N} \left(y_i * \log\left(\frac{y_i}{\hat{y}_i}\right) + (1 - y_i) * \log\left(\frac{1 - y_i}{1 - \hat{y}_i}\right)\right)$$
$$= \frac{1}{\mathrm{N}} \sum_{i=1}^{N} \left(y_i * \log\left(\frac{y_i}{\hat{y}_i}\right)\right) + \frac{1}{\mathrm{N}} \sum_{i=1}^{N} \left((1 - y_i) * \log\left(\frac{1 - y_i}{1 - \hat{y}_i}\right)\right)$$

So, now to find

$$\frac{\partial \frac{1}{\mathrm{N}} \sum_{i=1}^{N} \left(y_i * \log\left(\frac{y_i}{\hat{y}_i}\right)\right)}{\partial w_k} = y_i * \frac{1}{\mathrm{N}} \sum_{i=1}^{N} y_i * \frac{\partial \log\left(\frac{y_i}{\hat{y}_i}\right)}{\partial w_k}$$

We must apply the chain rule F'(w) = f'(g(w))g'(w) where g(w) = $\frac{y_i}{\hat{y}_i}$

Which requires us to find the derivative of g(w) = $\frac{y_i}{\hat{y}_i}$:

$$\frac{\partial \frac{y_i}{\hat{y}_i}}{\partial w_k} = \partial\, y_i * \left(1 + e^{-(w_0 + w_k X_{i,k})}\right) = y_i * \frac{\partial}{\partial w_k} e^{-(w_0 + w_k X_{i,k})} = -y_i * X_{i,k} * e^{-(w_0 + w_k X_{i,k})}.$$

Then, to find f', we find

$$\frac{\partial f\big(g(w_k)\big)}{\partial g(w_k)} = \frac{\partial}{\partial g(w_k)} \log\big(g(w_k)\big) = \frac{1}{g(w_k)} = \frac{\hat{y}_i}{y_i}$$

So, we combine these results to find

$$\frac{\partial \frac{1}{\mathrm{N}} \sum_{i=1}^{N} y_i * \left(\log\left(\frac{y_i}{\hat{y}_i}\right)\right)}{\partial w_k} = \frac{1}{\mathrm{N}} \sum_{i=1}^{N} y_i * \left(\frac{\hat{y}_i}{y_i}\right) * \left(-y_i * X_{i,k} * e^{-(w_0 + w_k X_{i,k})}\right)$$
$$= \frac{1}{\mathrm{N}} \sum_{i=1}^{N} \hat{y}_i * (-y_i) * X_{i,k} * e^{-(w_0 + w_k X_{i,k})}$$

Similarly, we can find the derivative for

$$\frac{\partial}{\partial w_k} \frac{1}{\mathrm{N}} \sum_{i=1}^{N} (1 - y_i) * \log\left(\frac{1 - y_i}{1 - \hat{y}_i}\right) = \frac{1}{\mathrm{N}} \sum_{i=1}^{N} (1 - y_i) * \frac{\partial}{\partial w_k} \log\left(\frac{1 - y_i}{1 - \hat{y}_i}\right)$$

by applying the chain rule F(w) = f'(g(w))*g'(w), where g(w) = $\frac{1 - y_i}{1 - \hat{y}_i}$, which we can rewrite for simplicity as

$$\mathrm{g}(w_k) = \frac{1-y_i}{1-\hat{y}_i} = \frac{1-y_i}{1-\dfrac{1}{1+e^{-(w_0+w_kX_{i,k})}}} = (1-y_i) * e^{(w_0+w_kX_{i,k})}$$

So,

$$\frac{\partial g(w_k)}{\partial w_k} = \frac{\partial (1-y_i) * e^{(w_0+w_kX_{i,k})}}{\partial w_k} = (1-y_i)\frac{\partial e^{(w_0+w_kX_{i,k})}}{\partial w_k} = X_{i,k} * (y_i - 1) * e^{(w_0+w_kX_{i,k})}$$

Then,

$$\frac{\partial f\big(g(w_k)\big)}{\partial g(w_k)} = \frac{\partial}{\partial g(w_k)}\log\big(g(w_k)\big) = \frac{1}{g(w_k)} = \frac{1-\hat{y}_i}{1-y_i} = \frac{1}{(1-y_i)(e^{(w_0+w_kX_{i,k})}+1)}$$

Which we then bring back to our original equation to solve

$$\frac{\partial}{\partial w_k}\frac{1}{\mathrm{N}}\sum_{i=1}^{N}(1-y_i) * \log\left(\frac{1-y_i}{1-\hat{y}_i}\right)$$
$$= \frac{1}{\mathrm{N}}\sum_{i=1}^{N}(y_i-1) * X_{i,j} * (y_i-1) * e^{(w_0+w_kX_{i,k})} * \frac{1}{(1-y_i)\big(e^{(w_0+w_kX_{i,k})}+1\big)}$$
$$= \frac{1}{\mathrm{N}}\sum_{i=1}^{N}X_{i,j} * (y_i-1) * e^{(w_0+w_kX_{i,k})} * \frac{1}{\big(e^{(w_0+w_kX_{i,k})}+1\big)}$$
$$= \frac{1}{\mathrm{N}}\sum_{i=1}^{N}X_{i,j} * (y_i-1) * \frac{e^{(w_0+w_kX_{i,k})}}{\big(e^{(w_0+w_kX_{i,k})}+1\big)}$$
$$= \frac{1}{\mathrm{N}}\sum_{i=1}^{N}X_{i,j} * (y_i-1) * \hat{y}_i$$

This just leaves us the last part of the original equation to derive, which is simple

$$\frac{\partial}{\partial w_k}\frac{1}{2}\lambda\sum_{j=1}^{4096}{w_j}^2 = \lambda * w_k$$

Now, to put that all together, we say

$$\frac{\partial \mathrm{L}(\vec{\mathrm{w}}, w_0, \lambda)}{\partial w_k} = \frac{1}{\mathrm{N}}\sum_{i=1}^{N}\hat{y}_i * (-y_i) * X_{i,k} * e^{-(w_0+w_kX_{i,k})} + \frac{1}{\mathrm{N}}\sum_{i=1}^{N}X_{i,j} * (y_i-1) * \hat{y}_i + \lambda * w_k$$
$$= \frac{1}{\mathrm{N}}\sum_{i=1}^{N}\big(\hat{y}_i * (-y_i) * X_{i,k} * e^{-(w_0+w_kX_{i,k})} + X_{i,j} * (y_i-1) * \hat{y}_i\big) + \lambda * w_k$$
$$= \frac{1}{\mathrm{N}}\sum_{i=1}^{N}X_{i,k} * \big(\hat{y}_i * (-y_i) * e^{-(w_0+w_kX_{i,k})} + (y_i-1) * \hat{y}_i\big) + \lambda * w_k$$

$$= \frac{1}{\mathrm{N}}\sum_{i=1}^{N} X_{i,k} * \left(\hat{y}_i - \hat{y}_i * y_i - \hat{y}_i * y_i * e^{-(w_0+w_k X_{i,k})}\right) + \lambda * w_k$$

$$= \frac{1}{\mathrm{N}}\sum_{i=1}^{N} X_{i,k} * \left(\hat{y}_i - \hat{y}_i * y_i(1 + e^{-(w_0+w_k X_{i,k})}\right) + \lambda * w_k$$

$$= \frac{1}{\mathrm{N}}\sum_{i=1}^{N} X_{i,k} * \left(\hat{y}_i - y_i(\frac{\hat{y}_i}{\hat{y}_i})\right) + \lambda * w_k$$

So, finally:

$$\frac{\partial \mathrm{L}(\vec{\mathrm{w}}, w_0, \lambda)}{\partial w_k} = \frac{1}{\mathrm{N}}\sum_{i=1}^{N}\left(X_{i,k} * (\hat{y}_i - y_i)\right) + \lambda * w_k$$

∂L/∂w0 derivation

With the loss function below, we will now similarly find the derivative with regards to w_0

$$\mathrm{L}(\vec{\mathrm{w}}, w_0, \lambda) = \frac{1}{\mathrm{N}}\sum_{i=1}^{N}\left(y_i * \log\left(\frac{y_i}{\hat{y}_i}\right) + (1-y_i) * \log\left(\frac{1-y_i}{1-\hat{y}_i}\right)\right) + \frac{1}{2}\lambda \sum_{j=1}^{4096} w_j{}^2$$

Again, we will start by taking the derivative as the sum of the derivatives of the two parts of the sum, starting with the first half:

$$\frac{\partial \frac{1}{\mathrm{N}}\sum_{i=1}^{N}\left(y_i * \log\left(\frac{y_i}{\hat{y}_i}\right) + (1-y_i) * \log\left(\frac{1-y_i}{1-\hat{y}_i}\right)\right)}{\partial w_0}$$

Which can be further split into two, allowing us to find the derivatives of each separately because

$$\frac{1}{\mathrm{N}}\sum_{i=1}^{N}\left(y_i * \log\left(\frac{y_i}{\hat{y}_i}\right) + (1-y_i) * \log\left(\frac{1-y_i}{1-\hat{y}_i}\right)\right)$$
$$= \frac{1}{\mathrm{N}}\sum_{i=1}^{N}\left(y_i * \log\left(\frac{y_i}{\hat{y}_i}\right)\right) + \frac{1}{\mathrm{N}}\sum_{i=1}^{N}\left((1-y_i) * \log\left(\frac{1-y_i}{1-\hat{y}_i}\right)\right)$$

So, now to find

$$\frac{\partial \frac{1}{\mathrm{N}}\sum_{i=1}^{N}\left(y_i * \log\left(\frac{y_i}{\hat{y}_i}\right)\right)}{\partial w_0} = y_i * \frac{1}{\mathrm{N}}\sum_{i=1}^{N} y_i * \frac{\partial \log\left(\frac{y_i}{\hat{y}_i}\right)}{\partial w_0}$$

We must apply the chain rule F'(w) = f'(g(w))g'(w) where g(w) = $\frac{y_i}{\hat{y}_i}$

Which requires us to find the derivative of g(w) = $\frac{y_i}{\hat{y}_i}$:

$\frac{\partial \frac{y_i}{\hat{y}_i}}{\partial w_0} = \partial\, y_i * \left(1 + e^{-(w_0 + w_k X_{i,k})}\right) = y_i * \frac{\partial}{\partial w_0} e^{-(w_0 + w_k X_{i,k})} = -y_i * e^{-(w_0 + w_k X_{i,k})}$.

Then, to find f', we find

$$\frac{\partial f\big(g(w_0)\big)}{\partial g(w_0)} = \frac{\partial}{\partial g(w_0)} \log\big(g(w_0)\big) = \frac{1}{g(w_0)} = \frac{\hat{y}_i}{y_i}$$

So, we combine these results to find

$$\frac{\partial \frac{1}{\mathrm{N}} \sum_{i=1}^{N} y_i * \left(\log\left(\frac{y_i}{\hat{y}_i}\right)\right)}{\partial w_0}$$

$$= \frac{1}{\mathrm{N}} \sum_{i=1}^{N} y_i * \left(\frac{\hat{y}_i}{y_i}\right) * \left(-y_i * e^{-(w_0 + w_k X_{i,k})}\right)$$

$$= \frac{1}{\mathrm{N}} \sum_{i=1}^{N} \hat{y}_i * (-y_i) * e^{-(w_0 + w_k X_{i,k})}$$

Similarly, we can find the derivative for

$$\frac{\partial}{\partial w_0} \frac{1}{\mathrm{N}} \sum_{i=1}^{N} (1 - y_i) * \log\left(\frac{1 - y_i}{1 - \hat{y}_i}\right) = \frac{1}{\mathrm{N}} \sum_{i=1}^{N} (1 - y_i) * \frac{\partial}{\partial w_0} \log\left(\frac{1 - y_i}{1 - \hat{y}_i}\right)$$

by applying the chain rule F(w) = f'(g(w))*g'(w), where g(w) = $\frac{1-y_i}{1-\hat{y}_i}$, which we can rewrite for simplicity as

$$\mathrm{g}(w_0) = \frac{1 - y_i}{1 - \hat{y}_i} = \frac{1 - y_i}{1 - \frac{1}{1 + e^{-(w_0 + w_k X_{i,k})}}} = (1 - y_i) * e^{(w_0 + w_k X_{i,k})}$$

So,

$$\frac{\partial g(w_0)}{\partial w_0} = \frac{\partial (1 - y_i) * e^{(w_0 + w_k X_{i,k})}}{\partial w_0} = (1 - y_i) \frac{\partial e^{(w_0 + w_k X_{i,k})}}{\partial w_0} = (y_i - 1) * e^{(w_0 + w_k X_{i,k})}$$

Then,

$$\frac{\partial f\big(g(w_0)\big)}{\partial g(w_0)} = \frac{\partial}{\partial g(w_0)} \log\big(g(w_0)\big) = \frac{1}{g(w_0)} = \frac{1 - \hat{y}_i}{1 - y_i} = \frac{1}{(1 - y_i)(e^{(w_0 + w_k X_{i,k})} + 1)}$$

Which we then bring back to our original equation to solve

$$\frac{\partial}{\partial w_0}\frac{1}{\mathrm{N}}\sum_{i=1}^{N}(1-y_i)*\log\left(\frac{1-y_i}{1-\hat{y}_i}\right)$$

$$=\frac{1}{\mathrm{N}}\sum_{i=1}^{N}(y_i-1)*(y_i-1)*e^{(w_0+w_kX_{i,k})}*\frac{1}{(1-y_i)\left(e^{(w_0+w_kX_{i,k})}+1\right)}$$

$$=\frac{1}{\mathrm{N}}\sum_{i=1}^{N}(y_i-1)*e^{(w_0+w_kX_{i,k})}*\frac{1}{\left(e^{(w_0+w_kX_{i,k})}+1\right)}$$

$$=\frac{1}{\mathrm{N}}\sum_{i=1}^{N}(y_i-1)*\frac{e^{(w_0+w_kX_{i,k})}}{\left(e^{(w_0+w_kX_{i,k})}+1\right)}$$

$$=\frac{1}{\mathrm{N}}\sum_{i=1}^{N}(y_i-1)*\hat{y}_i$$

This just leaves us the last part of the original equation to derive, which is even simpler this time

$$\frac{\partial}{\partial w_0}\frac{1}{2}\lambda\sum_{j=1}^{4096}{w_j}^2=0$$

Now, to put that all together, we say

$$\frac{\partial \mathrm{L}(\vec{\mathrm{w}},w_0,\lambda)}{\partial w_0}=\frac{1}{\mathrm{N}}\sum_{i=1}^{N}\hat{y}_i*(-y_i)*e^{-(w_0+w_kX_{i,k})}+\frac{1}{\mathrm{N}}\sum_{i=1}^{N}(y_i-1)*\hat{y}_i+0$$

$$=\frac{1}{\mathrm{N}}\sum_{i=1}^{N}\left(\hat{y}_i*(-y_i)*e^{-(w_0+w_kX_{i,k})}+(y_i-1)*\hat{y}_i\right)$$

$$=\frac{1}{\mathrm{N}}\sum_{i=1}^{N}\left(\hat{y}_i*(-y_i)*e^{-(w_0+w_kX_{i,k})}+(y_i-1)*\hat{y}_i\right)$$

$$=\frac{1}{\mathrm{N}}\sum_{i=1}^{N}\left(\hat{y}_i-\hat{y}_i*y_i-\hat{y}_i*y_i*e^{-(w_0+w_kX_{i,k})}\right)$$

$$=\frac{1}{\mathrm{N}}\sum_{i=1}^{N}\left(\hat{y}_i-\hat{y}_i*y_i(1+e^{-(w_0+w_kX_{i,k})}\right)$$

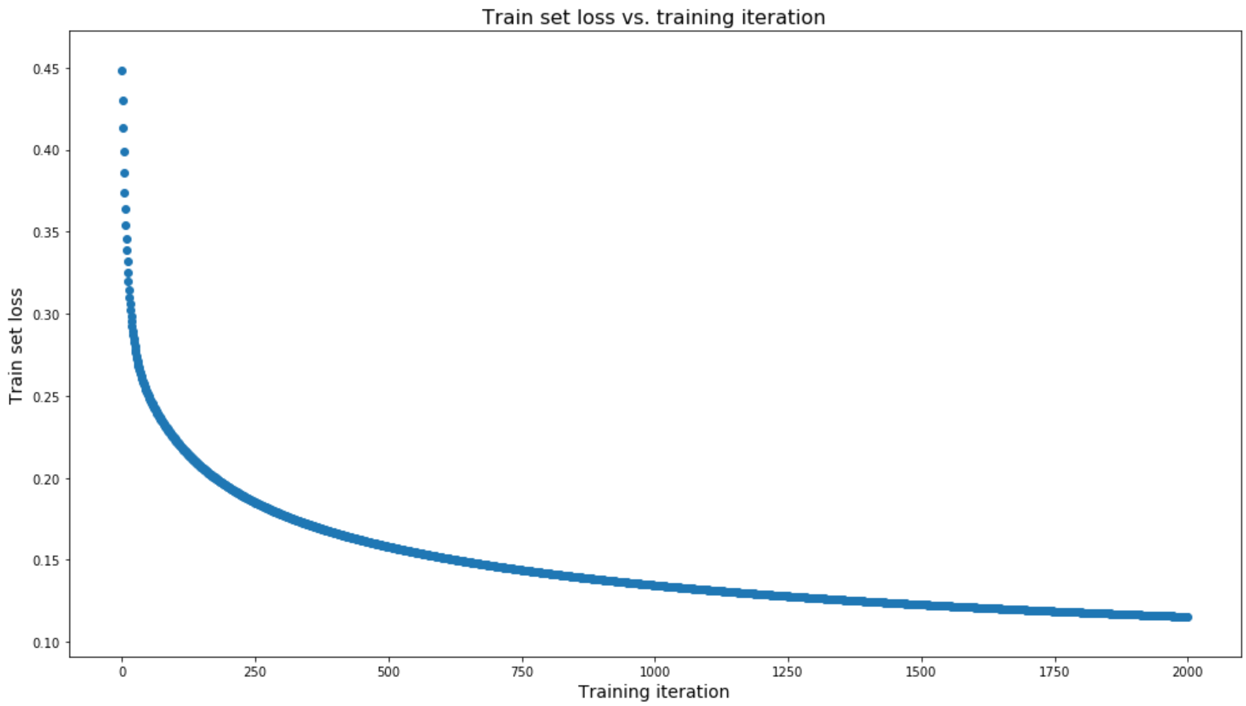
$$=\frac{1}{\mathrm{N}}\sum_{i=1}^{N}\left(\hat{y}_i-y_i(\frac{\hat{y}_i}{\hat{y}_i})\right)$$

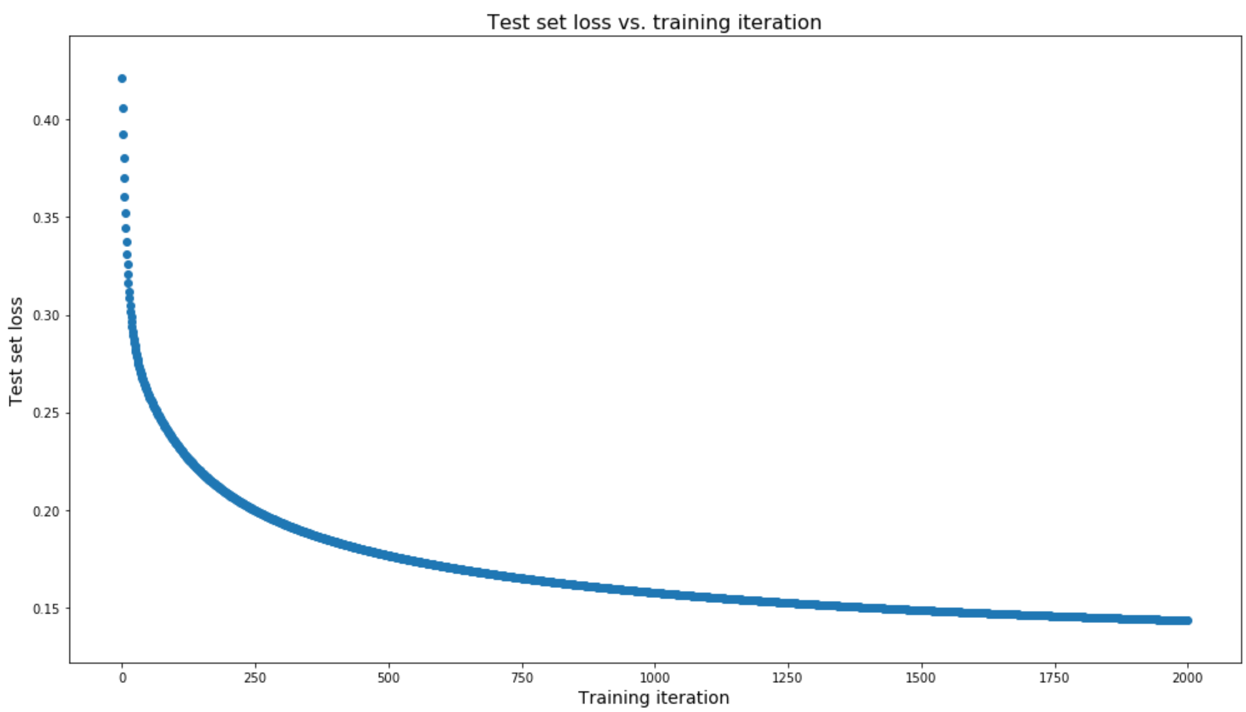
So, finally:

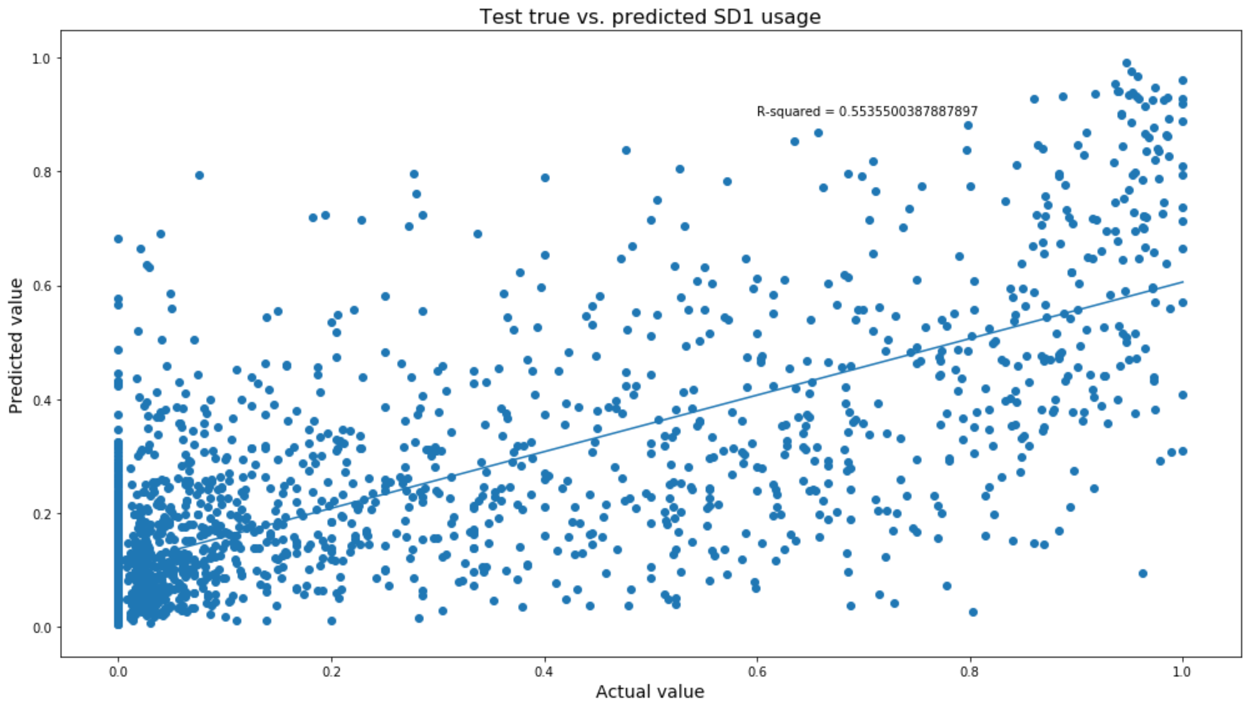
$$\frac{\partial \mathrm{L}(\vec{\mathrm{w}},w_0,\lambda)}{\partial w_0}=\frac{1}{\mathrm{N}}\sum_{i=1}^{N}(\hat{y}_i-y_i)$$

3. Gradient Descent Implementation

Results from the gradient descent implementation







Pearson R^2 value = 0.55355

The resulting error does not ever grow between iterations before reaching convergence within the λ = 0.0001 limitation, it is always converging.

10 most enhancing 6mers, with weights

1. GGGGGG = 0.7516851861509208
2. GGGAGG = 0.5557615127030084
3. GGAGGG = 0.4674845457590094
4. GGGTGG = 0.4607843124294966
5. GGTGGG = 0.443739283141239
6. GGGGGT = 0.43239217172185856
7. AGGGGG = 0.42510180949145804
8. GGGGAG = 0.40557013223742644
9. TGGGGG = 0.39704483654690115
10. GGGGGA = 0.3964454601591616

10 most repressive 6mers, with weights

1. GGTAAG = -0.4905625206301185
2. CGGTCG = -0.4526635880113442
3. CTTGGA = -0.37691215597236133
4. AGGTAA = -0.3247229520695178
5. CTTGGT = -0.3177554807667483
6. AGAAGA = -0.29436415586639525
7. AAGGTA = -0.28605283456665076
8. GAAGAT = -0.2829213533360671

9. CTTGGC = -0.2752746813347673
10. AAGAAG = -0.2711098799819469

Here we can see a pattern that the top 6mers to increase the odds ratio, or most enhancing motifs, are those with the most guanine (G), with a secondary preference for adenine (A). All are 4 or more G mixed with 1 or 2 A or T. A 6mer with lots of guanine appears to make a very good splice donor.

On the other hand, for most repressive 6mers, there is a pattern that they contain a 2mer of GG often paired with a 2mer of AA or TT. These 2mer pairs