1 Optimization

Introduce Langrangian equation, there are two constraints so we introduce two variables, α and β.

$$\begin{aligned} L(x_1, x_2, \alpha, \beta) = x_1^2 + x_2^2 - 1 - \alpha(x_1 + x_2 - 1) - \beta(x_1 - 2x_2) \\ s.t. \quad \alpha, \beta \geq 0 \end{aligned}$$

We need to solve for $\nabla L(x_1, x_2, \alpha, \beta) = 0$. It results in a system of four linear equations with four unknowns.

$$\begin{cases} \frac{\partial L(x_1,x_2,\alpha,\beta)}{\partial x_1} = 0 & \implies 2x_1 - \alpha - \beta = 0 \\ \frac{\partial L(x_1,x_2,\alpha,\beta)}{\partial x_2} = 0 & \implies 2x_2 - \alpha + 2\beta = 0 \\ \frac{\partial L(x_1,x_2,\alpha,\beta)}{\partial \alpha} = 0 & \implies -x_1 - x_2 = -1 \\ \frac{\partial L(x_1,x_2,\alpha,\beta)}{\partial \beta} = 0 & \implies -x_1 + 2x_2 = 0 \end{cases}$$

In turn, subtracting the fourth equation to the third one and viceversa, we obtain

$$x_1 = \frac{2}{3}$$
$$x_2 = \frac{1}{3}$$

Substituting these values for the first two equations we simplify the system to two linear equations such that:

$$\begin{cases} \alpha + \beta = \frac{4}{3} \\ \alpha - 2\beta = \frac{2}{3} \end{cases}$$

Hence,

$$\alpha = \frac{10}{9}$$
$$\beta = \frac{2}{9}$$

Having the values for x_1, x_2, α, β, we can check on the original equation and constraints described on the problem statement.

$$\begin{aligned} x_1 + x_2 - 1 = 0 \quad &\implies \frac{2}{3} + \frac{1}{3} - 1 = 0 \\ x_1 - 2x_2 \geq 0 \quad &\implies \frac{2}{3} - 2\frac{1}{3} \geq 0 \end{aligned}$$

We have obtained valid values for x_1 and x_2 hence we can now solve the original problem:

$$f(x_1, x_2) = \min_{x_1, x_2}(x_1^2 + x_2^2 - 1)$$
$$\therefore f(\frac{2}{3}, \frac{1}{3}) = (\frac{2}{3})^2 + (\frac{1}{3})^2 - 1 = -\frac{4}{9}$$

Collaborators:

Sources: Your Sources

2 Stochastic Process

Let e be expected number of tosses.

Suppose we toss it once. If get a tail (T) (probability $1/2$), then the expected number of tosses will be $e+1$. Suppose we toss it twice. If we get two heads (TT) (probability $1/4$), then the expected number of tosses will also be $e+1$[1] If we generalize for the sequence $HTT\dots T$ where $TT\dots T$ is a sequence of size k elements, table 1.

Outcome	Probability	Expected Value
T	$1/2$	$e+1$
HH	$1/4$	$e+1$
HTH	$1/8$	$e+2$
$\vdots$	$\vdots$	$\vdots$
$HTT\dots TH$	$1/2^{k+1}$	$e+k$
$HTT\dots TT$	$1/2^{k+1}$	$k+1$

Table 1: Caption

Thus, it follows:

$$e_k = \frac{1}{2}(e+1) + \frac{1}{4}(e+1) + \frac{1}{8}(e+2) + \ldots + \frac{1}{2^{k+1}}(e+k) + \frac{1}{2^{k+1}}(k+1)$$
$$e_k = e\Big(\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots + \frac{1}{2^{k+1}}\Big) + \frac{1}{2} + \frac{1}{2}\Big(\frac{1}{2} + \frac{2}{4} + \frac{3}{8} + \ldots + \frac{k}{2^k}\Big) + \frac{1}{2^{k+1}}(k+1)$$

The content of the first brackets is a geometric series with $r = 1/2$ and $a = 1/2$. We can compute the sum using the sum of the geometric series formula:

$$S_n = \frac{a(1-r^n)}{1-r}$$
$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots + \frac{1}{2^{k+1}} = \frac{1-(1/2)^{k+1}}{2(1-1/2)}$$
$$= 1-(1/2)^{k+1}$$

The content in the second brackets is an arithmetico-geometric sequence with $a_1 = 1$, $b_1 = 1/2$, $d = 1$ and $r = 1/2$. We can compute the sum using the following formula:

$$S_n = \frac{ab-(a+nd)br^n}{1-r} + \frac{dbr(1-r^n)}{(1-r)^2}$$

$$\frac{1}{2} + \frac{2}{4} + \frac{3}{8} + \ldots + \frac{k}{2^k} = \frac{(1/2)-(1+k)(1/2)(1/2)^k}{1-1/2} + \frac{(1/2)(1/2)(1-(1/2)^k)}{(1-1/2)^2}$$
$$= 2-(2^{-k})*(n+2)$$

Putting all together, we obtain

$$e_k = [e - e(1/2)^{k+1}] + \frac{1}{2}[2-(1/2)^k(k+2)] + \frac{1}{2} + (1/2)^{k+1}(k+1)$$
$$e_k = 2^{k+1} * \Big(\frac{3}{2} - (1/2)^{k+1}(k+2) + (1/2)^{k+1}(k+1)\Big)$$

[1] This is because the first element of the sequence is a head, so we would already have it.

Hence, the expected value in terms of k:

$$e_k = 3 * 2^k - 1$$

3 SVM

Using Langrangian equation, there are four constraints so we introduce four variables, α, β, λ and γ.

$$L(w,b,\xi,\hat{\xi}) = \frac{1}{2}||w||^2 + C\sum_{i=1}^{N}\xi_i + C\sum_{i=1}^{N}\hat{\xi}_i - \sum_{i=1}^{N}\alpha_i(w^Tx_i + b + \epsilon + \xi_i - y_i) \tag{1}$$

$$- \sum_{i=1}^{N}\beta_i(w^Tx_i + b - \epsilon - \hat{\xi}_i - y_i) \tag{2}$$

$$- \sum_{i=1}^{N}\lambda_i\xi_i - \sum_{i=1}^{N}\gamma_i\hat{\xi}_i \tag{3}$$

$$s.t.\ \alpha,\beta,\lambda,\gamma \geq 0 \tag{4}$$

We need to solve for $\nabla L(w,b,\epsilon,\xi,\alpha,\beta,\lambda,\gamma) = 0$. It results in a system of linear equations.

$$\frac{\partial L}{\partial w} = 0 \implies w - \sum_{i=1}^{N}\alpha_i x_i - \sum_{i=1}^{N}\beta_i x_i = 0 \ \therefore\ \hat{w} = \sum_{i=1}^{N}\alpha_i x_i + \sum_{i=1}^{N}\beta_i x_i \tag{5}$$

$$\frac{\partial L}{\partial b} = 0 \implies -\sum_{i=1}^{N}\alpha_i - \sum_{i=1}^{N}\beta_i = 0 \ \therefore\ \sum_{i=1}^{N}\alpha_i = -\sum_{i=1}^{N}\beta_i \tag{6}$$

$$\frac{\partial L}{\partial \xi} = 0 \implies C - \sum_{i=1}^{N}\alpha_i - \sum_{i=1}^{N}\lambda_i = 0 \therefore C - \sum_{i=1}^{N}\alpha_i = \sum_{i=1}^{N}\lambda_i \tag{7}$$

$$\frac{\partial L}{\partial \hat{\xi}} = 0 \implies C + \sum_{i=1}^{N}\beta_i - \sum_{i=1}^{N}\gamma_i = 0 \therefore C + \sum_{i=1}^{N}\beta_i = \sum_{i=1}^{N}\gamma_i \tag{8}$$

Notice that the equality from equation 6 can be used to replace for α in equation 5, such that

$$\hat{w} = \sum_{i=1}^{N}\alpha_i x_i + \sum_{i=1}^{N}\beta_i x_i = \sum_{i=1}^{N}\alpha_i x_i - \sum_{i=1}^{N}\alpha_i x_i = 0 \tag{9}$$

We proceed to replace these variables according to equations 6,7,8 and 9 into the lagrangian equation 1 so as to rewrite the equation in terms of α.

$$\begin{aligned}L(\hat{w}, b, \xi, \hat{\xi}, \alpha) = \frac{1}{2} * ||0||^2 + C\sum_{i=1}^{N} \xi_i + C\sum_{i=1}^{N} \hat{\xi}_i - \sum_{i=1}^{N} \alpha_i x_i \cdot (\sum_{j=1}^{N} \alpha_j x_j - \sum_{j=1}^{N} \alpha_j x_j) \\ - b * \sum_{i=1}^{N} \alpha_i - \epsilon * \sum_{i=1}^{N} \alpha_i - \sum_{i=1}^{N} \alpha_i \xi_i + \sum_{i=1}^{N} \alpha_i y_i \\ + \sum_{i=1}^{N} \alpha_i x_i \cdot (\sum_{j=1}^{N} \alpha_j x_j - \sum_{j=1}^{N} \alpha_j x_j) \\ + b * \sum_{i=1}^{N} \alpha_i - \epsilon * \sum_{i=1}^{N} \alpha_i - \sum_{i=1}^{N} \alpha_i \hat{\xi}_i - \sum_{i=1}^{N} \alpha_i y_i \\ - C * \sum_{i=1}^{N} \xi_i + \sum_{i=1}^{N} \alpha_i \xi_i - C * \sum_{i=1}^{N} \hat{\xi}_i + \sum_{i=1}^{N} \alpha_i \hat{\xi}_i\end{aligned}$$

Simplifying this, we obtain:

$$L(\hat{w}, b, \xi, \hat{\xi}, \alpha) = -2\epsilon * \sum_{i-1}^{N} \alpha_i$$

Hence, the dual problem for the proposed linear SVM becomes:

$$\hat{\alpha} = \arg \max_{0 \leq \alpha \leq C} \Big(-2\epsilon * \sum_{i-1}^{N} \alpha_i \Big)$$

4 Neural Network

Best configuration:

Epoch	Learning Rate	Activation	Reg.	Reg. Rate	Layers	Neurons	Output
147	0.01	ReLU	L2	0.01	2	(3,3)	0.465

Table 2: Caption

The learning rate appears to affect the speed at which the algorithm converges to the minima. Too low would require an increasing number of epochs, while too high it might cause it to jump optimal solutions into alternative local minima. Adding L2 regularization appeared to yield positive results as long as it did not go above 0.03. Any value higher than this negatvely affected performance. In general, using ReLU appears to yield better results than either tanh or the sigmoid function for the given dataset. Optimal performance was obtained when using two hidden layers (with two features. Increasing the number of hidden layers, for this particular dataset, decreased the performance of the model.

5 Logistic Regression

I first formatted the data into 124 columns (target + features), indicating whether a feature was active with 1s and inactive with 0s. In addition, the target (y_i) was modified to hold values such that $y_i \in 0, 1$.

Problems encountered include dealing with the singularity matrix (not invertible), but was solved by adding small regularizing parameters. I also had to train it in batches as the diagonal matrix would become too large. Upon implementing all this steps, I was able to obtain an accuracy of 100%, yet I do not fully understand why, converging almost immediately.

The following is the procedure to obtain the gradient and the hessian.

$$\nabla L(w) = \sum_{i=1}^{N} (y_i x_i - x_i)\alpha_i, \quad where \quad \alpha = \frac{e^{w^T x_i}}{1 + e^{w^T x_i}}$$
$$= \sum_{i=1}^{N} x_i (y_i - \alpha_i) = X * (y - \alpha)$$

$$H^{-1} = \nabla^2 L(w) = -\sum_{i=1}^{N} x_i x_i^T \alpha_i (1 - \alpha_i) = -XRX^T$$

Weights were updated accordingly using Newton's Method:

$$w_{t+1} = w_t - H^1 \cdot \nabla L(w)$$