**ElasticNet: Una técnica intermedia:**

Hasta el momento hemos podido ver dos técnicas de regularización en las cuales añadimos un componente de penalización en el proceso donde encontramos los valores de los parámetros 𝛽 minimizando la función de error.

Por ejemplo, si usamos el método de Mínimos Cuadrados Ordinarios, tenemos por definición nuestra función definida como:

https://lh5.googleusercontent.com/0ZuL5aZCMAw31jMCuC1jAatO9NVTLIpFUL5yqXJdXH6du9Qboru3-rT66Q-O0VE1ehK5k9pk61JU-emm2eK1CqZmEyja6RuaSvuDmZQvN2VNRyFmm5ON4RYCQlETlv1YVDMEgpoH

Ahora bien. Si aplicamos la regularización L1 también conocida como Lasso (Least Absolute Shrinkage and Selection Operator), tenemos una ecuación de la forma:

https://lh3.googleusercontent.com/5PDYLjW19y0AzHm6WodPpgNYhv4TNTXLxI02QZcouVSZfKItlFurZdbSdYy9mgTlZjuNgCjidSAR_tMa1pAX5onCqS-G26-NjM7Ywtz4X9wlXhg1S821z4bTN7IdyA7smD7cxK7l

donde tenemos un parámetro de ajuste llamado ƛ que si tiene valores altos para el problema mandará el valor de 𝛽j a 0.

Por otro lado. Si aplicamos la regularización L2 también conocida como Ridge, tendremos la siguiente ecuación:

https://lh6.googleusercontent.com/FoTb5M6vY2gE3PLzbYcsKQDuzEz2m37p4IOztejOwnfZe2WXVScA0unEkIpPV1xOreYu-3tTxQL9cXG8zboRX1F8sVCN66AheZi-0jD4iWn-xy4vpglVBsMc9T7gxFEzZDkpLmhj

Tendremos una penalización también pero que no tiene la posibilidad de llevar los valores de los coeficientes a cero. Sin embargo esto nos permitirá realizar el intercambio de +sesgo por -varianza.

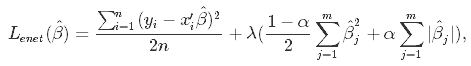
Recordando que :

1. Ninguna de las dos es mejor que la otra para todos los casos.
2. Lasso envía algunos coeficientes a cero permitiendo así seleccionar variables significativas para el modelo.
3. Lasso funciona mejor si tenemos pocos predictores que influyen sobre el modelo.
4. Ridge funciona mejor si es el caso contrario y tenemos una gran cantidad.

Para aplicarlos y decidir cuál es el mejor en la práctica, podemos probar usando alguna técnica como cross-validation iterativamente. o bien, podemos combinarlos…

Regularización ElasticNet

Es común encontrarnos en la literatura con un camino intermedio llamado ElasticNet. Esta técnica consiste en combinar las dos penalizaciones anteriores en una sola función. Así, nuestra ecuación de optimización quedará:



Donde tenemos ahora un parámetro adicional 𝛂 que tiene un rango de valores entre 0 y 1. Si 𝛂 = 0 , ElasticNet se comportará como Ridge, y si 𝛂 = 1 , se comportará como Lasso. Por lo tanto, nos brinda todo el espectro lineal de posibles combinaciones entre estos dos extremos.

1. Tenemos una forma de probar ambas L1 y L2 al tiempo sin perder información.
2. Supera las limitaciones individuales de ellas.
3. Si hace falta experiencia, o el conocimiento matemático de fondo, puede ser la opción preferente para probar la regularización.

ElasticNet con Scikit-learn

Para implementar esta técnica añadimos primero el algoritmo ubicado en el módulo linear\_model.

from sklearn.linear\_model import ElasticNet

Y luego simplemente lo inicializamos con el constructor ElasticNet() y entrenamos con la función fit().

regr = ElasticNet(random\_state=0)

regr.fit(X, y)