

Nama	: Adhimas Aryo Bimo
NIM	: 13523052
Model	: ANN

1. ANN bekerja dengan prinsip dasar feedforward neural network + backpropagation. Alur kerja ANN sederhana adalah sebagai berikut :
 - a. Init parameter
 - Bobot W diinisialisasi dengan metode tertentu, dalam hal ini saya mengimplementasikan xavier, he, dan zeros.
 - Bias diinisialisasi dengan 0
 - Layer diatur sesuai dengan layer_sizer dan activations
 - b. Forward pass
 - Input X masuk ke layer pertama
 - Tiap layer akan dihitung nilai Z, pada hal ini Z merupakan pre komputasi untuk menghitung nilai aktivasi nantinya. Z dihitung melalui rumus berikut:

$$Z = A \cdot W + B$$
 - c. Hitung Loss
 - Untuk regresi perhitungan loss dalam hal ini menggunakan mse_loss dalam fungsi yang saya definisikan. Namun nilai loss ini bisa di-adjust pada pengembangan lebih lanjut.
 - d. Backward pass
 - Setelah loss dihitung, dihitung gradien turunan loss terhadap output dA
 - Lalu, dilakukan propagasi mundur ke setiap layer, secara matematis dapat ditulis sebagai berikut:

$$dz = dA \cdot f'(Z)$$
 - Lalu, gradien bobot dan bias dihitung dalam fungsi berikut

$$\nabla W = \frac{A_{prev}^T \cdot dZ}{N}, \quad \nabla b = \frac{\sum dZ}{N}$$
 - Update dA untuk layer sebelumnya dengan $dA = dZ \cdot W^T$
 - e. Update parameter
 - Bobot dan bias diperbarui dengan gradient descent

$$W = W - \eta \cdot \nabla W, \quad b = b - \eta \cdot \nabla b$$
 - Jika ada regulasi L1/L2, maka akan ditambahkan pada gradien untuk generalisasi
 - f. Looping deh per epoch
 - Data tiap epoch akan diacak
 - Mini-batch diproses, loss dihitung, backprop dijalankan
 - Ketika semua batch udah selesau, lanjut ke epoch berikutnya

2. Ada perbedaan ANN yang saya buat dengan MLPRegressor dari scikit-learn. Model bawaan tersebut lebih bagus karena udah ada adaptive optimizer, learning rate scheduling, dll. Optimizer yang saya buat juga hanya vanilla gradient descent sedangkan pada bawaan scikit-learn sudah ada Adam pada optimizernya.
3. Mungkin bisa ditambahkan hidden layer didalamnya agar lebih banyak node koneksi untuk melihat pola dalam data. Selain itu model bisa dioptimasi dengan menambahkan Adam atau Momentum SGD agar ketika gradient descent bisa update weight dan biasanya lebih baik.