

Marcellinus Geofani Sihaloho

MLP

1. Jika menggunakan model MLP dengan 3 hidden layer (256-128-64) menghasilkan underfitting pada dataset ini, modifikasi apa yang akan dilakukan pada arsitektur? Jelaskan alasan setiap perubahan dengan mempertimbangkan bias-variance tradeoff!

Modifikasi Arsitektur (Underfitting):

- Tambah neuron/layer (misal 512–256–128–64) → untuk menurunkan bias.
- Ganti aktivasi (misal LeakyReLU) → hindari dead neuron.
- Kurangi regularisasi/dropout jika terlalu ketat.
- Bias-variance tradeoff: naikkan kompleksitas → kurangi bias, jaga agar variance tetap stabil.

2. Selain MSE, loss function apa yang mungkin cocok untuk dataset ini? Bandingkan kelebihan dan kekurangannya, serta situasi spesifik di mana alternatif tersebut lebih unggul daripada MSE!

- Huber Loss
Kelebihan : Gabungan MSE (untuk error kecil) dan MAE (untuk error besar) → robust & smooth
Kekurangan : Perlu pilih delta threshold
Situasi unggul : Ada outlier tapi tetap ingin gradient smooth
- MAE (Mean Absolute Error)
Kelebihan : Lebih tahan terhadap outlier
Kekurangan : Tidak smooth (gradient tidak kontinu di 0) → lebih lambat konvergensi
Situasi unggul : Data memiliki banyak outlier

3. Jika salah satu fitur memiliki range nilai 0-1, sedangkan fitur lain 100-1000, bagaimana ini memengaruhi pelatihan MLP? Jelaskan mekanisme matematis (e.g., gradien, weight update) yang terdampak!

- Gradien akan tidak seimbang → bobot terkait fitur besar akan mengalami update lebih cepat, sedangkan fitur kecil hampir tidak diubah.
- Vanishing/Exploding Gradient lebih mungkin terjadi pada layer awal.

Mekanisme matematis:

Backpropagation menggunakan turunan loss terhadap bobot ($\partial L / \partial w = \partial L / \partial z * \partial z / \partial w$), dan jika input x (fitur) besar, maka $\partial z / \partial w = x$ juga besar, menyebabkan update Δw menjadi terlalu agresif.

Solusi:

Gunakan StandardScaler atau MinMaxScaler untuk menormalkan semua fitur agar berada pada skala yang seragam → membantu konvergensi yang stabil dan lebih cepat.

4. Tanpa mengetahui nama fitur, bagaimana Anda mengukur kontribusi relatif setiap fitur terhadap prediksi model? Jelaskan metode teknikal (e.g., permutation importance, weight analysis) dan keterbatasannya!

- Permutation Importance
Ukur penurunan performa (misal R^2) saat satu fitur diacak
Keterbatasan : Lambat, tidak cocok untuk fitur saling berkorelasi
- Weight Analysis
Amati bobot layer pertama → fitur dengan bobot besar punya pengaruh lebih kuat
Keterbatasan : Tidak bisa menilai non-linearitas atau interaksi antar fitur

5. Bagaimana Anda mendesain eksperimen untuk memilih learning rate dan batch size secara optimal? Sertakan analisis tradeoff antara komputasi dan stabilitas pelatihan!

- Coba kombinasi (grid/random search).
- LR tinggi: cepat tapi bisa divergen. Rendah: stabil tapi lambat.
- Batch kecil: stabil, generalisasi baik. Besar: cepat, risiko overfit.
- Gunakan early stopping & visualisasi loss.