

Analisa
MLP (Multilayer Perceptron) – Tabular Data
RegresiUTSTelkom.csv
Elieser Pasaribu – 1103223209

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!

Jawab:

Untuk mengatasi underfitting dengan arsitektur MLP 3 hidden layer(256-128-64):

- Tingkatkan kapasitas model:
 - Tambahkan lebih banyak neuron ke layer yang ada (misalnya, 512-25-128)
 - Tambahkan layer tambahan untuk hierarki fitur yang lebih kompleks
 - Kurangi tingkat dropout jika terlalu agresif (di bawah 0,2)
- Tingkatkan representasi fitur:
 - Tingkatkan rekayasa fitur (fitur polinomial, istilah interaksi)
 - Gunakan transformasi data yang lebih canggih (logaritma, akar kuadrat, pangkat)
 - Pertimbangkan pengurangan dimensi untuk menghilangkan noise
- Modifikasi regularisasi:
 - Kurangi tingkat dropout
 - Kurangi regularisasi bobot (L1/L2) jika ada
 - Sesuaikan parameter BatchNormalization
- Penyesuaian pelatihan:
 - Tingkatkan jumlah epoch pelatihan
 - Bereksperimen dengan learning rate
 - Coba optimizers yang berbeda (RMSprop, SGD dengan momentum)

Dari perspektif tradeoff bias-variance, underfitting menunjukkan bias yang tinggi. Model terlalu sederhana untuk menangkap pola yang mendasari. Setiap modifikasi

yang disarankan bertujuan untuk mengurangi bias sambil memantau kinerja validasi dengan hati-hati untuk menghindari peningkatan varians (overfitting).

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!

Jawab:

Fungsi loss alternatif selain MSE:

- Mean Absolute Error (MAE):
 - Lebih robust terhadap outlier dibandingkan MSE
 - Kelebihan: Kurang dipengaruhi oleh nilai ekstrem, menghasilkan gradien yang lebih stabil
 - Terbaik ketika: Data mengandung outlier, prediksi median lebih diinginkan daripada rata-rata
- Huber Loss:
 - Menggabungkan manfaat MSE dan MAE
 - Kelebihan: Tahan terhadap outlier sambil mempertahankan efisiensi pelatihan MSE untuk kesalahan yang lebih kecil
 - Terbaik ketika: Dataset memiliki beberapa outlier tetapi Anda masih ingin perilaku seperti MSE untuk sebagian besar sampel
- Log-Cosh Loss:
 - Mendekati MSE untuk kesalahan kecil, MAE untuk kesalahan besar
 - Kelebihan: Turunan yang halus di mana saja, tahan terhadap outlier
 - Terbaik ketika: Membutuhkan fungsi loss dengan sifat numerik yang baik
- MSLE (Mean Squared Logarithmic Error):
 - Memberikan penalti lebih pada perkiraan yang terlalu rendah dibandingkan yang terlalu tinggi
 - Kelebihan: Baik untuk distribusi target yang miring ke kanan
 - Terbaik ketika: Kesalahan relatif lebih penting daripada kesalahan absolut

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!

Jawab:

Pengaruh perbedaan skala fitur (0-1 vs 100-1000):

- Fitur dengan range besar akan mendominasi fungsi loss dan update gradient
- Secara matematis, jika w_1 adalah weight untuk fitur skala kecil dan w_2 untuk fitur skala besar, partial derivative terhadap w_2 akan jauh lebih besar: $\partial L / \partial w_2 \gg \partial L / \partial w_1$
- Ini menyebabkan weight untuk fitur skala besar diupdate lebih agresif
- Learning rate optimal menjadi sulit ditemukan: terlalu kecil untuk fitur skala kecil, terlalu besar untuk fitur skala besar

- Solusinya adalah normalisasi/standarisasi fitur sehingga semua fitur memiliki kontribusi seimbang dalam proses pembelajaran
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!

Jawab:

- Permutation Importance: Mengukur penurunan performa ketika nilai fitur diacak
 - Partial Dependence Plots: Menunjukkan hubungan fitur-target setelah mengontrol fitur lain
 - SHAP (SHapley Additive exPlanations): Memberikan kontribusi setiap fitur berdasarkan teori game
 - Integrated Gradients: Mengukur kontribusi fitur berdasarkan gradien terintegrasi
 - Analisis bobot untuk layer pertama: Melihat magnitude bobot sebagai indikator kasar
 - Keterbatasan: Permutation importance mengasumsikan fitur independen; analisis bobot tidak menangkap interaksi kompleks; metode berbasis gradien sensitif terhadap noise lokal.
5. Bagaimana Anda mendesain eksperimen untuk memilih learning rate dan batch size secara optimal? Sertakan analisis tradeoff antara komputasi dan stabilitas pelatihan!

Jawab:

Eksperimen untuk memilih learning rate dan batch size:

- Learning rate: Gunakan learning rate scheduler seperti cyclical learning rate atau learning rate range test untuk menemukan LR optimal. Mulai dari nilai kecil ($1e-5$) dan tingkatkan secara bertahap sampai loss meningkat tajam.
- Batch size: Uji beberapa nilai (8, 16, 32, 64, 128) dan evaluasi trade-off: -
Batch kecil: Update lebih sering, noise lebih besar, dapat keluar dari local minima, tapi komputasi lebih lambat per epoch
- Batch besar: Update lebih stabil, komputasi lebih cepat per epoch, tapi bisa terjebak di local minima Trade-off komputasi vs stabilitas:
- Batch besar + LR kecil = komputasi efisien, konvergensi lambat tapi stabil
- Batch kecil + LR besar = konvergensi potensial lebih cepat, tapi tidak stabil
- Solusi optimal: Batch menengah dengan LR schedule adaptif