

## Tugas 2 Pemelajaran Mesin Lanjut

**Nama : Marshal Arijona Sinaga**

**NPM : 2006560983**

### Keterangan:

- seluruh proses komputasi dilakukan pada *environment cloud* FloydHub dengan menggunakan GPU Tesla V-80
- Dataset: CIFAR10 → 50000 train, 10000 test
- Label pada *data set* CIFAR-10 dikodekan sebagai berikut:
  - 0: airplane
  - 1: automobile
  - 2: bird
  - 3: cat
  - 4: deer
  - 5: dog
  - 6: frog
  - 7: horse
  - 8: ship
  - 9: truck
- Hyperparameter terdiri dari jumlah epoch, train\_batch, tiny\_error dan optimizer. Pada Tugas ini, *hyperparameter* yang di-tuning adalah jumlah epoch dan optimizer.

1. Lakukan Klasifikasi untuk dataset cifar menggunakan Fully Connected Neural Networks dengan arsitektur dasar sebagai berikut:

- Fully Connected Layer (1000 hidden node)
- Activation Layer (ReLU function)
- Fully Connected Layer (1000 hidden node)
- Activation Layer (ReLU function)
- Output Layer (10 kelas)
- Activation Layer (softmax function)
- Classification Result Fitur yang digunakan bebas (handcrafted/non-convolution).

Hitung metrik (akurasi, sensitivity , specificity, dan f1-score) untuk data train dan data test.

### Jawab:

- *Training phase* dilakukan menggunakan optimizer *stochastic gradient descent* dengan nilai momentum sebesar 0.9 dan nilai *learning rate* sebesar 0.001.
- *Loss function* yang digunakan berupa *cross-entropy*.
- Epoch: 50
- *Train batch size*: 200 → *mini batch size* yang digunakan pada *training data* ketika training phase.
- *Test batch size*: 10000 → *mini batch size* yang digunakan pada *testing data* ketika testing phase
- *Full train batch size*: 50000 → *mini batch size* yang digunakan pada *training data* ketika testing phase

- Fitur yang digunakan adalah  $32 * 32$  piksel CIFAR-10 yang sudah ditransformasi menjadi *grayscale*.
- Citra *grayscale* terlebih dahulu dinormalisasi dengan memilih nilai *mean* = 0.5 dan *SD* = 0.5

### **Metrik test-set**

Accuracy: 0.2967

Sensitivity untuk setiap kelas:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.3087	0.2855	0.0	0.0	0.2832	0.2919	0.2749	0.2445	0.3269	0.3512

Average sensitivity = 0.23668

Specificity untuk setiap kelas:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.9268	0.9249	0.9	0.9	0.9195	0.9281	0.9206	0.9256	0.9398	0.9403

Average specificity = 0.92256

f1-score untuk setiap kelas:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.3285	0.3095	0.0	0.0	0.278	0.3275	0.2822	0.2938	0.3915	0.4077

average f1-score = 0.26187

### **Metrik train-set**

Accuracy: 0.2918

Sensitivity untuk setiap kelas:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.299	0.2855	0.0	0.0	0.2683	0.2948	0.2625	0.2467	0.3189	0.3472

Average sensitivity = 0.23229

Specificity untuk setiap kelas:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.9244	0.9252	0.9	0.9	0.9179	0.9292	0.9198	0.9257	0.9368	0.9411

Average specificity = 0.92201

f1-score untuk setiap kelas:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.3124	0.3112	0.0	0.0	0.2629	0.3337	0.2733	0.2954	0.3764	0.408

Average f1-score = 0.25733

2. Anda dapat memperbaiki hasil yang diperoleh dengan melakukan 3(tiga) skenario eksperimen, yaitu: feature engineering, menambahkan layer, dan hyperparameter tuning (pada classifier).

**Feature engineering:**

Pada model ini, fitur yang digunakan adalah  $3 * 32 * 32$  piksel CIFAR. Selanjutnya data dinormalisasi dengan memilih *mean* sebesar (0.5, 0.5, 0.5) dan SD sebesar (0.5, 0.5, 0.5) (citra RGB). Untuk parameter yang lain diatur sama dengan model awal.

**metrik test-set:**

Accuracy: 0.3928

Sensitivity untuk setiap kelas:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.4651	0.4085	0.0	0.0	0.3489	0.3455	0.3417	0.3696	0.4559	0.4442

Average sensitivity = 0.31794

Specificity untuk setiap kelas:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.9394	0.9423	0.9	0.9	0.9339	0.9327	0.954	0.9366	0.9533	0.9435

Average specificity = 0.93357

f1-score untuk setiap kelas:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.459	0.4468	0.0	0.0	0.3803	0.3733	0.4416	0.4026	0.5158	0.4696

Average f1-score = 0.3489

**metrik train-set**

Accuracy: 0.3986

Sensitivity untuk setiap kelas:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.4587	0.4115	0.0	0.0	0.3357	0.359	0.3453	0.3788	0.4818	0.4534

Average sensitivity = 0.32242

Specificity untuk setiap kelas:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.09375	0.944	0.9	0.9	0.9321	0.9454	0.9539	0.9374	0.9556	0.9465

Average specificity = 0.850865

f1-score untuk setiap kelas:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.4463	0.4551	0.0	0.0	0.3662	0.3916	0.444	0.4104	0.5392	0.4876

Average f1-score = 0.35404

### **Adding more layers:**

Pada model ini, ditambahkan 1 *layer fully connected* dengan jumlah neuron sebanyak 500 buah. Untuk parameter lain diatur sama dengan model yang pertama kali digunakan.

### **Metrik test-set:**

Accuracy: 0.2627

Sensitivity untuk setiap kelas:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.275	0.2329	0.0	0.0	0.2338	0.2379	0.2272	0.0	0.3178	0.2899

Average sensitivity = 0.18145

Specificity untuk setiap kelas:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.9297	0.9178	0.9	0.9	0.9175	0.9291	0.9162	0.9	0.9397	0.9418

Average specificity = 0.91918

f1-score untuk setiap kelas:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
---------	---------	---------	---------	---------	---------	---------	---------	---------	---------

0.3256	0.2522	0.0	0.0	0.2506	0.3022	0.2407	0.0	0.3855	0.373
--------	--------	-----	-----	--------	--------	--------	-----	--------	-------

Average f1-score = 0.21298

### **metrik train-set**

Accuracy: 0.2573

Sensitivity untuk setiap kelas:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.2569	0.2352	0.0	0.0	0.2153	0.2456	0.2244	0.0	0.3047	0.2941

Average sensitivity = 0.17762

Specificity untuk setiap kelas:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.9251	0.9188	0.9	0.9	0.9156	0.9311	0.9165	0.9	0.9355	0.9426

Average specificity = 0.91852

f1-score untuk setiap kelas:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.2979	0.2587	0.0	0.0	0.2342	0.3132	0.242	0.0	0.3634	0.3781

Average f1-score = 0.20875

### **Hyperparameter tuning**

Pada model ini nilai EPOCH diubah menjadi 100. Selain itu, optimotor diubah menggunakan Adaptive optimization (Adam) dengan nilai *learning rate* 0.0002. Untuk parameter yang lain, disamakan dengan model yang pertama kali digunakan.

### **metrik test-set**

Accuracy: 0.461

Sensitivity untuk setiap kelas:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.537	0.5602	0.388	0.318	0.3918	0.3713	0.4454	0.5496	0.5633	0.4871

Average sensitivity = 0.46117

Specificity untuk setiap kelas:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.9413	0.9449	0.9315	0.9191	0.9313	0.9355	0.9443	0.946	0.9548	0.953

Average specificity = 0.94017

f1-score untuk setiap kelas:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class
0.4979	0.5273	0.3855	0.2837	0.3858	0.3985	0.4738	0.529	0.5792	0.5324

average f1-score = 0.45931

### **metrik train-test**

Accuracy: 0.7356

Sensitivity untuk setiap kelas:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.7912	0.8082	0.6951	0.7097	0.6953	0.6738	0.7089	0.7913	0.7644	0.7255

Average sensitivity = 0.73634

Specificity untuk setiap kelas:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.9699	0.9781	0.962	0.9606	0.9657	0.9663	0.9755	0.9733	0.9767	0.9784

Average specificity = 0.97065

f1-score untuk setiap kelas:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.7576	0.8056	0.6749	0.6739	0.6931	0.6856	0.7437	0.7745	0.7776	0.7647

Average f1-score = 0.73512

3. Analisis hasil yang anda dapatkan untuk masing-masing skenario.

### **Jawab:**

feature engineering:

- Akurasi dengan melakukan feature engineering meningkat (train dan test)
- Average sensitivity dengan melakukan feature engineering meningkat (train dan test)
- Average specificity pada test set meningkat sementara pada train set menurun jika menggunakan feature engineering
- Average f1 score dengan melakukan feature engineering meningkat (train dan test)

- Peningkatan nilai akurasi diduga karena fitur RGB yang digunakan untuk menggantikan fitur grayscale lebih mampu untuk mendeskripsi statistik data untuk setiap kelas sehingga model yang dihasilkan lebih baik dalam mengenali input yang diberikan
- Nilai sensitivity model yang menggunakan feature engineering pada test cukup rendah, sehingga mengurangi confidence prediksi yang dihasilkan oleh model.
- Nilai akurasi test set dan train set tidak jauh berbeda, sehingga dapat dikatakan bahwa model dengan menggunakan feature engineering tidak mengalami overfitting.

Adding more layers:

- Akurasi model yang ditambahkan layer menurun (test dan train)
- Average sensitivity model yang ditambahkan layer menurun (train dan test)
- Average specificity model yang ditambahkan layer menurun (train dan test)
- Average f1 score model yang ditambahkan layer menurun (train dan test)
- Model yang digunakan di awal sudah mengalami underfitting karena nilai akurasi pada saat training memang sudah rendah. Artinya kapasitas model awal memang masih kurang. Terdapat dua kemungkinan yang menyebabkannya yaitu kurangnya epoch atau kurangnya kapasitas model (kurang layer)
- skenario penambahan layer menyebabkan metrik pengukuran model semakin menurun. Artinya, kemungkinan besar model membutuhkan epoch yang lebih besar agar mencapai titik yang lebih mendekati konvergen sehingga metrik pengukurannya lebih baik

Hyperparameter tuning:

- Akurasi model yang dituning meningkat dibandingkan dengan model awal (train dan test)
- Average sensitivity model yang dituning meningkat dibandingkan dengan model awal (train dan test)
- Average specificity model yang dituning meningkat dibandingkan dengan model awal (train dan test)
- Average f1-score model yang dituning meningkat dibandingkan dengan model awal (train dan test)
- Nilai akurasi train yang lebih tinggi dibandingkan dengan test menunjukkan bahwa kemungkinan model mengalami overfitting
- Nilai sensitivity rendah namun specificity tinggi. Hal ini mengurangi confidence dari prediksi yang dihasilkan oleh model.

## **B. Convolutional Neural Networks**

1. Lakukan Klasifikasi untuk dataset CIFAR-10 menggunakan Convolutional Neural Networks dengan arsitektur dasar sebagai berikut:

- Convolutional Layer (25 maps, kernel 3x3)
- Pooling Layer (2x2)
- Convolutional Layer (50 maps, kernel 3x3)
- Pooling Layer (2x2)
- Hidden Layer (100 neuron)
- Activation Function (ReLU function)
- Output Layer (10 kelas)
- Activation Function (Softmax function)
- Classification Result

- *Training phase* dilakukan menggunakan optimotor *stochastic gradient descent* dengan nilai momentum sebesar 0.9 dan nilai *learning rate* sebesar 0.001.
- *Loss function* yang digunakan berupa *cross-entropy*.
- Epoch: 50
- *Train batch size*: 200 → *mini batch size* yang digunakan pada *training data* ketika training phase.
- *Test batch size*: 10000 → *mini batch size* yang digunakan pada *testing data* ketika *testing phase*
- *Full train batch size*: 50000 → *mini batch size* yang digunakan pada *training data* ketika *testing phase*
- Fitur yang digunakan adalah 3 \* 32 \* 32 piksel CIFAR-10
- Citra *grayscale* terlebih dahulu dinormalisasi dengan memilih nilai *mean* = (0.5, 0.5, 0.5) dan *SD* = (0.5, 0.5, 0.5) (RGB)

2. Hitung metrik (akurasi, sensitivity , specificity, dan f1-score) untuk untuk data train dan data test.

### **Metrik test-set**

Accuracy: 0.4918

Sensitivity untuk setiap kelas:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.5746	0.5776	0.3829	0.3467	0.4418	0.4268	0.5004	0.5419	0.5654	0.5231

Average sensitivity = 0.44812

Specificity untuk setiap kelas:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.9449	0.9541	0.9304	0.9236	0.935	0.9321	0.9566	0.9466	0.9608	0.9525

Average specificity = 0.94366

f1-score untuk setiap kelas:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.533	0.5828	0.3768	0.3234	0.4253	0.4032	0.5538	0.5292	0.606	0.5487

Average f1-score = 0.48822

### **Metrik train-test**

Accuracy: 0.5147

Sensitivity for each class:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
---------	---------	---------	---------	---------	---------	---------	---------	---------	---------



0.5924	0.5866	0.4074	0.3841	0.4449	0.4618	0.5292	0.5564	0.6059	0.5447
--------	--------	--------	--------	--------	--------	--------	--------	--------	--------

Average sensitivity = 0.51134

Specificity for each class:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.948	0.9572	0.9325	0.9279	0.9371	0.936	0.9563	0.9482	0.963	0.9556

Average specificity = 0.94618

f1-score for each class:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.5574	0.6012	0.3984	0.3629	0.4382	0.4386	0.5682	0.5436	0.6366	0.5736

Average f1-score = 0.51187

3. Anda dapat memperbaiki hasil yang didapat dengan melakukan 2 (dua) skenario eksperimen, yaitu menambah layer dan hyperparameter tuning (pada classifier)

**Jawab:**

#### **Adding more layers**

Pada model ini, ditambahkan 1 *layer convolutional* dengan ukuran kernel sebesar 3x3 sebanyak 50 buah, 1 *layer max pooling* dengan ukuran kernel sebesar 2x2 dan 1 *layer fully connected* dengan jumlah neuron sebanyak 50 buah. Untuk *hyperparameter* yang lain disamakan dengan model CNN yang pertama kali digunakan.

#### **Metrik test-set**

Accuracy: 0.5644

Sensitivity for each class:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.5992	0.6865	0.5014	0.3926	0.4941	0.4756	0.6257	0.5766	0.6834	0.5945

Average sensitivity = 0.56296

Specificity for each class:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.9588	0.9621	0.9298	0.9354	0.945	0.9392	0.9643	0.9571	0.9645	0.9608

Average specificity = 0.9517

f1-score for each class:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.6147	0.6714	0.4102	0.4077	0.5	0.4619	0.6526	0.5961	0.6817	0.6215

Average f1-score = 0.56178

### **Metrik train-set**

Accuracy: 0.6351

Sensitivity for each class:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.6572	0.7341	0.5862	0.4867	0.5367	0.5795	0.6801	0.6518	0.7633	0.6706

Average sensitivity = 0.63462

Specificity for each class:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.9643	0.9718	0.9375	0.9471	0.9524	0.9479	0.9685	0.965	0.9717	0.9691

Average specificity = 0.95953

f1-score for each class:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.6683	0.7405	0.489	0.5069	0.5554	0.5516	0.6986	0.6689	0.7537	0.6967

Average f1-score = 0.63296

### **Hyperparameter tuning**

Pada model yang di hyperparameter tuning nilai EPOCH ditambah dari 50 menjadi 100  
Selain itu untuk optimotor diganti menjadi optimotor Adaptive optimization (Adam) dengan nilai *learning rate* sebesar 0.0002.

### **Metrik test-set**

Accuracy: 0.657

Sensitivity for each class:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.6704	0.7357	0.5827	0.5036	0.5807	0.5618	0.7295	0.7245	0.8093	0.6743

Average sensitivity = 0.65725

Specificity for each class:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
---------	---------	---------	---------	---------	---------	---------	---------	---------	---------

0.9693	0.9747	0.9439	0.9436	0.9588	0.9529	0.9701	0.9678	0.9707	0.9674
--------	--------	--------	--------	--------	--------	--------	--------	--------	--------

Average specificity = 0.96192

f1-score for each class:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.6971	0.7544	0.53	0.4972	0.6057	0.5693	0.7303	0.7172	0.7698	0.6907

Average f1-score = 0.65617

### **Metrik train-set**

Accuracy: 0.8446

Sensitivity for each class:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.8541	0.8832	0.8173	0.7674	0.8135	0.8041	0.8586	0.8728	0.9016	0.871

Average sensitivity = 0.84436

Specificity for each class:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.9843	0.9905	0.9714	0.976	0.982	0.9755	0.9874	0.9851	0.9878	0.9876

Average specificity = 0.98276

f1-score for each class:

class 0	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9
0.8563	0.8987	0.7766	0.7759	0.8258	0.7911	0.8727	0.8692	0.896	0.8796

Average f1-score = 0.84419

4. Analisis hasil yang anda dapatkan untuk masing-masing skenario.

### **Jawab:**

penambahan layer:

- penambahan layer meningkatkan nilai akurasi dari model (train dan test)
  - Nilai akurasi pada train set lebih besar sekitar 10% dibandingkan dengan test set.
- Kemungkinan besar model mengalami *overfitting*.
- Nilai akurasi pada test set masih belum maksimal (di kisaran 50%)
  - Average sensitivity pada model yang ditambahkan layernya meningkat (train dan test)
  - Average specificity pada model yang ditambahkan layernya meningkat +- 0.01 (train dan test)
  - Average f1-score pada model yang ditambahkan layernya meningkat (train dan test).
  - Peningkatan akurasi dengan menambahkan layer diperkirakan mampu mendapatkan fitur yang lebih deskriptif dari input (higher level features) sehingga mampu membedakan kelas input dengan lebih baik.
  - pada test set, model memiliki sensitivitas tergolong sedang (+- 0.5) dan spesifitas tinggi.

- pada train set, model memiliki sensitivitas tergolong sedang ( $\pm 0.6$ ) dan spesifitas tinggi.
- Model yang baik haruslah memiliki spesifitas dan sensitivitas yang tinggi.

parameter tuning:

- Akurasi model yang dituning lebih tinggi dibandingkan model awal (train dan test)
- Average sensitivity model yang dituning lebih tinggi dibandingkan model awal (train dan test)
- Average specificity model yang dituning lebih tinggi dibandingkan model awal (train dan test)
- f1 score model yang dituning lebih tinggi dibandingkan model awal (train dan test)
- penggunaan epoch yang lebih tinggi memungkinkan model mendekati titik yang lebih konvergen. Selain itu, penggantian optimatior dari stochastic gradient descent menuju Adam diduga menyebabkan proses training menjadi lebih stabil sehingga lebih cepat menuju titik konvergen.
- Kapasitas model masih belum ideal dikarenakan nilai sensitivitas yang masih kurang tinggi (meskipun nilai spesifitasnya sangat tinggi. Model yang baik harus memiliki f1-score, sensitivitas dan spesifitas yang tinggi)
- Nilai akurasi train set jauh lebih tinggi dibandingkan test set. Kemungkinan model mengalami overfitting.