

כווןון פרמטרים היפר-פרמטריים לרשת מחוברת מלאה על MNIST

1. ארכיטקטורת המודל

קונפיגורציה בסיסית

1. שכבת קלט: מפשטת את תמונות ה- 28×28 פיקסלים לוקטור בגודל 784.
2. ארבע שכבות Dense:
 - שכבה 1: 82 נירונים
 - שכבה 2: 4 נירונים
 - שכבה 3: 76 נירונים
 - שכבה 4: 12 נירונים
3. Batch Normalization ו-ReLU: כל שכבה מלווה ב-Batch Normalization, ולאחר מכן הפעלת ReLU.
4. Dropout: חלה הסתברות dropout מסוימת לאחר כל הפעלת activation כדי להפחית overfitting.
5. שכבת פלט: שכבת Dense אחרונה עם 10 יחידות (אחת לכל קטגוריית ספרה), תוך שימוש בהפעלה מסוג softmax.

- Batch Normalization מסייעת לייצב ולהאיץ את האימון על ידי הפחתת שינויים פנימיים בקורלציות (internal covariate shift).
- Dropout פועל כצורת רגולריזציה על ידי "כיבוי" רנדומלי של חלק מהנירונים בכל מעגל, ובכך מונעת תלות יתר בנירון בודד.
- הפעלת ReLU היא בחירה פופולרית וסטנדרטית לרשתות מחוברות מלאה בזכות פשטותה ויעילותה.

2. תוצאות

לאחר 5 ניסויים, נמצא הסט האופטימלי של פרמטרים:

- שיעור הלמידה: 0.00694
- שיעור Dropout (שכבה 1): 0.1
- שיעור Dropout (שכבה 2): 0.1
- שיעור Dropout (שכבה 3): 0.1
- שיעור Dropout (שכבה 4): 0.2

הגדרה זו השיגה דיוק אימות של כ-96.2%. כאשר המודל אומן מחדש על כל מערכת האימון (באמצעות אותו פיצול), הושג דיוק מבחן של כ-95.9%.

Trial 26 Complete [00h 00m 25s]
val_accuracy: 0.9370999932289124

Best val_accuracy So Far: 0.9402999877929688
Total elapsed time: 00h 09m 35s

Search: Running Trial #27

Value	Best Value So Far	Hyperparameter
0.5	0.1	dropout_0
0.5	0.2	dropout_1
0.2	0.5	dropout_2
0.3	0.1	dropout_3
0.0001	0.01	learning_rate
10	10	tuner/epochs
0	4	tuner/initial_epoch
0	1	tuner/bracket
0	1	tuner/round

Trial 28 Complete [00h 00m 30s]
val_accuracy: 0.821399986743927

Best val_accuracy So Far: 0.9402999877929688
Total elapsed time: 00h 10m 46s

Search: Running Trial #29

Value	Best Value So Far	Hyperparameter
0.1	0.1	dropout_0
0.5	0.2	dropout_1
0.5	0.5	dropout_2
0.3	0.1	dropout_3
0.01	0.01	learning_rate
10	10	tuner/epochs
0	4	tuner/initial_epoch
0	1	tuner/bracket
0	1	tuner/round

Trial 16 Complete [00h 00m 16s]
val_accuracy: 0.7768999934196472

Best val_accuracy So Far: 0.9157999753952026
Total elapsed time: 00h 05m 24s

Search: Running Trial #17

Value	Best Value So Far	Hyperparameter
0.5	0.5	dropout_0
0.1	0.1	dropout_1
0.3	0.3	dropout_2
0.3	0.3	dropout_3
0.01	0.01	learning_rate
10	4	tuner/epochs
4	2	tuner/initial_epoch
2	2	tuner/bracket
2	1	tuner/round
0013	0008	tuner/trial_id

Trial 18 Complete [00h 00m 24s]
val_accuracy: 0.9330999851226807

Best val_accuracy So Far: 0.9330999851226807
Total elapsed time: 00h 06m 13s

Search: Running Trial #19

Value	Best Value So Far	Hyperparameter
0.3	0.5	dropout_0
0.2	0.1	dropout_1
0.5	0.2	dropout_2
0.4	0.2	dropout_3
0.001	0.001	learning_rate
4	10	tuner/epochs
0	4	tuner/initial_epoch
1	2	tuner/bracket
0	2	tuner/round

Trial 19 Complete [00h 00m 21s]
val_accuracy: 0.7559999823570251

Best val_accuracy So Far: 0.9330999851226807
Total elapsed time: 00h 06m 34s

Search: Running Trial #20

Value	Best Value So Far	Hyperparameter
0.1	0.5	dropout_0
0.2	0.1	dropout_1
0.1	0.2	dropout_2
0.1	0.2	dropout_3
0.01	0.001	learning_rate
4	10	tuner/epochs
0	4	tuner/initial_epoch
1	2	tuner/bracket
0	2	tuner/round

Trial 10 Complete [00h 00m 17s]
val_accuracy: 0.7825000286102295

Best val_accuracy So Far: 0.9045000076293945
Total elapsed time: 00h 03m 34s

Search: Running Trial #11

Value	Best Value So Far	Hyperparameter
0.2	0.5	dropout_0
0.5	0.1	dropout_1
0.3	0.2	dropout_2
0.3	0.2	dropout_3
0.0001	0.001	learning_rate
2	2	tuner/epochs
0	0	tuner/initial_epoch
2	2	tuner/bracket
0	0	tuner/round

Trial 11 Complete [00h 00m 25s]
val_accuracy: 0.5608000159263611

Best val_accuracy So Far: 0.9045000076293945
Total elapsed time: 00h 03m 59s

Search: Running Trial #12

Value	Best Value So Far	Hyperparameter
0.5	0.5	dropout_0
0.4	0.1	dropout_1
0.2	0.2	dropout_2
0.5	0.2	dropout_3
0.01	0.001	learning_rate
2	2	tuner/epochs
0	0	tuner/initial_epoch
2	2	tuner/bracket
0	0	tuner/round

Trial 13 Complete [00h 00m 15s]
val_accuracy: 0.9156000018119812

Best val_accuracy So Far: 0.9156000018119812
Total elapsed time: 00h 04m 40s

Search: Running Trial #14

Value	Best Value So Far	Hyperparameter
0.5	0.5	dropout_0
0.1	0.1	dropout_1
0.3	0.2	dropout_2
0.3	0.2	dropout_3
0.01	0.001	learning_rate
4	4	tuner/epochs
2	2	tuner/initial_epoch
2	2	tuner/bracket
1	1	tuner/round
0008	0001	tuner/trial_id

Trial 9 Complete [00h 00m 16s]
val_accuracy: 0.8970000147819519

Best val_accuracy So Far: 0.9045000076293945
Total elapsed time: 00h 03m 17s

Search: Running Trial #10

Value	Best Value So Far	Hyperparameter
0.4	0.5	dropout_0
0.2	0.1	dropout_1
0.5	0.2	dropout_2
0.4	0.2	dropout_3
0.001	0.001	learning_rate
2	2	tuner/epochs
0	0	tuner/initial_epoch
2	2	tuner/bracket
0	0	tuner/round

Trial 1 Complete [00h 00m 49s]
val_accuracy: 0.486299991607666

Best val_accuracy So Far: 0.486299991607666
Total elapsed time: 00h 00m 49s

Search: Running Trial #2

Value	Best Value So Far	Hyperparameter
0.5	0.5	dropout_0
0.1	0.4	dropout_1
0.2	0.3	dropout_2
0.2	0.5	dropout_3
0.001	0.0001	learning_rate
2	2	tuner/epochs
0	0	tuner/initial_epoch
2	2	tuner/bracket
0	0	tuner/round

Trial 5 Complete [00h 00m 16s]
val_accuracy: 0.604200005531311

Best val_accuracy So Far: 0.9045000076293945
Total elapsed time: 00h 01m 53s

Search: Running Trial #6

Value	Best Value So Far	Hyperparameter
0.2	0.5	dropout_0
0.3	0.1	dropout_1
0.4	0.2	dropout_2
0.1	0.2	dropout_3
0.001	0.001	learning_rate
2	2	tuner/epochs
0	0	tuner/initial_epoch
2	2	tuner/bracket
0	0	tuner/round

Trial 6 Complete [00h 00m 26s]
val_accuracy: 0.7505000233650208

Best val_accuracy So Far: 0.9045000076293945
Total elapsed time: 00h 02m 19s

Search: Running Trial #7

Value	Best Value So Far	Hyperparameter
0.2	0.5	dropout_0
0.2	0.1	dropout_1
0.4	0.2	dropout_2
0.2	0.2	dropout_3
0.0001	0.001	learning_rate
2	2	tuner/epochs
0	0	tuner/initial_epoch
2	2	tuner/bracket
0	0	tuner/round

```

Trial 8 Complete [00h 00m 25s]
val_accuracy: 0.5485000014305115

Best val_accuracy So Far: 0.9045000076293945
Total elapsed time: 00h 03m 00s

Search: Running Trial #9

Value      |Best Value So Far|Hyperparameter
0.5         |0.5              |dropout_0
0.1         |0.1              |dropout_1
0.3         |0.2              |dropout_2
0.3         |0.2              |dropout_3
0.01        |0.001            |learning_rate
2           |2                |tuner/epochs
0           |0                |tuner/initial_epoch
2           |2                |tuner/bracket
0           |0                |tuner/round

```

```

Trial 7 Complete [00h 00m 16s]
val_accuracy: 0.5688999891281128

Best val_accuracy So Far: 0.9045000076293945
Total elapsed time: 00h 02m 35s

Search: Running Trial #8

Value      |Best Value So Far|Hyperparameter
0.5         |0.5              |dropout_0
0.4         |0.1              |dropout_1
0.2         |0.2              |dropout_2
0.3         |0.2              |dropout_3
0.0001      |0.001            |learning_rate
2           |2                |tuner/epochs
0           |0                |tuner/initial_epoch
2           |2                |tuner/bracket
0           |0                |tuner/round

```

ההליך החיפוש עם Hyperband

אז איך Hyperband באמת עובד? מדובר באלגוריתם שיודע לבדוק מלא תצורות של היפר-פרמטרים בצורה יעילה. הוא בוחר כמה שיותר שילובים, אבל לא מבזבז זמן על כאלה שלא נראים מבטיחים. במקום לתת לכל תצורה לרוץ עד הסוף, הוא מפסיק את אלה שפחות מצליחות כבר בשלבים הראשונים.

כמה ניסיונות זה לקח?

בפרויקט הזה, Hyperband בדק 30 ניסיונות בסך הכול. כל ניסיון ייצג שילוב אחר של שיעורי דרופאאוט לשכבות ושיעור למידה. ההליך כולו לקח 11 דקות. כן, זה כל מה שהוא היה צריך כדי למצוא את התצורה הכי טובה! אלגוריתם כזה ממש חוסך זמן כי הוא לא מתעכב על תצורות שלא עובדות טוב.

מה הוא מצא?

• שיעורי דרופאאוט:

- שכבה 1: 0.1
- שכבה 2: 0.1
- שכבה 3: 0.1
- שכבה 4: 0.2

• שיעור למידה: 0.01

התוצאות של המודל הכי טוב

אחרי ש-Hyperband מצא את ההיפר-פרמטרים הכי טובים, המודל אומן מחדש, וזה מה שהוא השיג:

• דיוק באימות: 0.95

• דיוק בבדיקה: 0.94

למה Hyperband כל כך יעיל?

היופי ב-Hyperband זה שהוא יודע לעצור ניסיונות שלא הולכים לשום מקום מוקדם, ולתת יותר זמן ומשאבים לתצורות שנראות מבטיחות. זה אומר שאתה מקבל תוצאות מצוינות בזמן קצר מאוד. זה בדיוק מה שעשינו פה – תוך 30 ניסיונות בלבד, קיבלנו מודל עם ביצועים מרשימים על MNIST.

```

Best val_accuracy So Far: 0.9528999924659729
Total elapsed time: 00h 11m 01s
Best dropout rates: [0.1, 0.1, 0.4, 0.1]
Best learning rate: 0.001

```