

Double-click (or enter) to edit

You need to work on a popular Fashion MNIST dataset for this HW. The dataset includes tiny images of fashion pieces. The objective is to create a set of supervised learning models that can predict the type of item based on its image.

In order to load the dataset you need to have tensorflow V2 on your computer. Use the following code to install the package

```
1 # !pip install --upgrade tensorflow
```

You can also check the version of it using the following code.

```
1 import tensorflow as tf
2 tf.__version__
```

```
↗ '2.19.0'
```

Now, it's time to load the dataset

```
1 from tensorflow import keras
2 fashion_mnist = keras.datasets.fashion_mnist
3 (X_train, y_train), (X_test, y_test) = fashion_mnist.load_data()
```

```
↗ Downloading data from https://storage.googleapis.com/tensorflow/tf-keras-datasets/train-labels-idx1-ubyte.gz
29515/29515 ————— 0s 0us/step
Downloading data from https://storage.googleapis.com/tensorflow/tf-keras-datasets/train-images-idx3-ubyte.gz
26421880/26421880 ————— 0s 0us/step
Downloading data from https://storage.googleapis.com/tensorflow/tf-keras-datasets/t10k-labels-idx1-ubyte.gz
5148/5148 ————— 0s 0us/step
Downloading data from https://storage.googleapis.com/tensorflow/tf-keras-datasets/t10k-images-idx3-ubyte.gz
4422102/4422102 ————— 0s 0us/step
```

As can be seen from the above code, the dataset was divided into train and test sets. Let's take a look at the X_train

```
1 X_train.shape
```

```
↗ (60000, 28, 28)
```

As it is clear, the train dataset (X_train) contains 60,000 images of size 28 x 28. We can visualize one of the images using the following code:

```
1 import matplotlib as mpl
2 import matplotlib.pyplot as plt
3 %matplotlib inline
4
5 sample_image = X_train[10]
6 plt.imshow(sample_image, cmap='binary')
7 plt.axis('off')
8 plt.show()
```

```
...
```

The `y_train` also includes values between 0 and 9. Each represents a particular category. For example, we can check the value of `y_train` for the above image.

```
1 y_train[10]
...
```

The above code shows that the image belongs to category 0. To get the associated label with each category, you can use the following code:

```
1 class_names = ['T-shirt/top', 'Trouser', 'Pullover', 'Dress', 'Coat', 'Sandal', 'Shirt', 'Sneaker', 'Bag', 'Ankle boot']
2 print(class_names[y_train[10]])
...

1 Start coding or generate with AI.
```

- **Task1:** Use the train set to train various supervised models and evaluate their performance using the test set.
 - Use different supervised learning models.
 - Use different metrics such as **accuracy**, **precision**, **AUC**, and ... in your model evaluation.
 - It is not enough to report the metrics. It is crucial that you interpret the metrics for each model and compare them across different models.
 - You may need to use the cross validation methods for hyperparameter selection.
 - Specify the model that outperforms the other models.
- **Task2:** Use the best model to predict your own fashion pieces.
 - Take a picture of ten fashion pieces of your own (take pictures in square format).
 - Resize images to the correct size (28,28).
 - Grayscale your images.
 - Visualize all the images side by side
 - Use the best model in Task 1 to predict the label of each of your own images.
 - How accurate is the final result?

✓ Output

- Make sure to put descriptive comments on your code
- Use the markdown cell format in Jupiter to add your own interpretation to the result in each section.
- Make sure to keep the output of your runs when you want to save the final version of the file.
- The final work should be very well structured and should have a consistent flow of analysis.

✓ 0. Setup

```
1 # %pip -q install -U tensorflow scikit-learn matplotlib seaborn pillow
2
3 import sys, platform, numpy as np, matplotlib.pyplot as plt, seaborn as sns
4 import tensorflow as tf
5 from tensorflow import keras
6 from tensorflow.keras import layers
7 from sklearn.metrics import accuracy_score, precision_score, classification_report, confusion_matrix, roc_auc_score
8 from sklearn.model_selection import train_test_split, GridSearchCV, StratifiedKFold
9 from sklearn.preprocessing import StandardScaler
10 from sklearn.pipeline import Pipeline
11 from sklearn.linear_model import LogisticRegression
12 from sklearn.svm import SVC
13 from sklearn.ensemble import RandomForestClassifier
14 from PIL import Image
15 import os, glob
16
17 print("Python:", sys.version)
18 print("Platform:", platform.platform())
19 print("TensorFlow:", tf.__version__)
...
```

✓ 1. Load Fashion-MNIST and Basic EDA

```

1 # Load Fashion-MNIST
2 (X_train, y_train), (X_test, y_test) = keras.datasets.fashion_mnist.load_data()
3
4 X_train.shape, X_test.shape, y_train.shape, y_test.shape
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```

✓ 2. Preprocessing

We'll prepare two pipelines:

- For classical ML (LogReg/SVM/RF): flatten 28x28 to 784 features; scale if needed.
- For CNN: keep 28x28x1, normalize to.

```

1 # Normalization
2 X_train_cnn = (X_train.astype("float32") / 255.0)[..., np.newaxis]
3 X_test_cnn = (X_test.astype("float32") / 255.0)[..., np.newaxis]
4
5 # Flatten for classical ML
6 X_train_flat = X_train.reshape(len(X_train), -1).astype("float32")/255.0
7 X_test_flat = X_test.reshape(len(X_test), -1).astype("float32")/255.0
8
9 num_classes = 10

```

✓ 3. Baselines and Classical Models (Task 1)

We will evaluate:

- Logistic Regression (multinomial)
- Linear SVM and RBF SVM (using probabilities via probability=True)
- Random Forest

We'll compute accuracy, macro-precision, and OvR ROC-AUC (requires predict_proba). For SVM linear we'll do two runs: one with calibrated probs via probability=True (note the slower training) or use decision_function with OneVsRestClassifier+CalibratedClassifierCV if desired. To keep runtime sane, we'll limit grid sizes.

```

1 def evaluate_model(clf, X_tr, y_tr, X_te, y_te, model_name):
2     y_proba = None
3     if hasattr(clf, "predict_proba"):
4         y_proba = clf.predict_proba(X_te)
5     elif hasattr(clf, "decision_function"):

```

```

6     # Convert decision scores to probabilities via softmax-like normalization per class
7     # Note: not a calibrated probability; acceptable for a rough AUC comparison if needed
8     scores = clf.decision_function(X_te)
9     # If binary shape, convert to two-column
10    if scores.ndim == 1:
11        scores = np.vstack([-scores, scores]).T
12    # Min-max per row to 0..1
13    s_min = scores.min(axis=1, keepdims=True)
14    s_max = scores.max(axis=1, keepdims=True)
15    denom = np.where((s_max - s_min)==0, 1, (s_max - s_min))
16    y_proba = (scores - s_min) / denom
17
18    y_pred = clf.predict(X_te)
19
20    acc = accuracy_score(y_te, y_pred)
21    prec_macro = precision_score(y_te, y_pred, average='macro', zero_division=0)
22
23    auc_ovr = None
24    if y_proba is not None:
25        # One-vs-Rest multiclass AUC with macro averaging
26        y_true_bin = keras.utils.to_categorical(y_te, num_classes=num_classes)
27        try:
28            auc_ovr = roc_auc_score(y_true_bin, y_proba, average='macro', multi_class='ovr')
29        except Exception:
30            auc_ovr = None
31
32    print(f"{model_name} - Accuracy: {acc:.4f} | Macro-Precision: {prec_macro:.4f} | OvR AUC: {auc_ovr if auc_ovr is not None else 'None'}")
33    return {"model": model_name, "acc": acc, "prec_macro": prec_macro, "auc_ovr": auc_ovr}
34
35 results = []

1 results

...

```

✓ 3.1 Logistic Regression (with small grid search)

```

1 pipe_lr = Pipeline([
2     ("scaler", StandardScaler(with_mean=False)), # with_mean=False safer for sparse-like; here dense but fine
3     ("clf", LogisticRegression(max_iter=200, multi_class='multinomial', solver='saga', n_jobs=-1))
4 ])
5
6 param_lr = {
7     "clf__C": [1.0, 0.5, 2.0]
8 }
9
10 cv = StratifiedKFold(n_splits=3, shuffle=True, random_state=42)
11 grid_lr = GridSearchCV(pipe_lr, param_lr, cv=cv, scoring='accuracy', n_jobs=-1, verbose=0)
12 grid_lr.fit(X_train_flat, y_train)
13 best_lr = grid_lr.best_estimator_
14 print("LR best params:", grid_lr.best_params_)
15
16 res_lr = evaluate_model(best_lr, X_train_flat, y_train, X_test_flat, y_test, "LogisticRegression")
17 results.append(res_lr)

...

```

✓ 3.2 Support Vector Machines

Linear SVM (probability=True uses Platt scaling internally; slower):

```

1 pipe_svm_lin = Pipeline([
2     ("scaler", StandardScaler(with_mean=False)),
3     ("clf", SVC(kernel='linear', probability=True))
4 ])
5
6 param_svm_lin = {"clf__C": [0.5, 1.0]}
7 grid_svm_lin = GridSearchCV(pipe_svm_lin, param_svm_lin, cv=cv, scoring='accuracy', n_jobs=-1, verbose=0)
8 grid_svm_lin.fit(X_train_flat, y_train)
9 best_svm_lin = grid_svm_lin.best_estimator_
10 print("SVM-linear best params:", grid_svm_lin.best_params_)
11

```

```

12 res_svm_lin = evaluate_model(best_svm_lin, X_train_flat, y_train, X_test_flat, y_test, "SVM-Linear")
13 results.append(res_svm_lin)
14
...

```

✓ RBF SVM:

```

1 pipe_svm_rbf = Pipeline([
2     ("scaler", StandardScaler(with_mean=False)),
3     ("clf", SVC(kernel='rbf', probability=True))
4 ])
5
6 param_svm_rbf = {"clf__C": [1.0], "clf__gamma": ['scale']}
7 grid_svm_rbf = GridSearchCV(pipe_svm_rbf, param_svm_rbf, cv=cv, scoring='accuracy', n_jobs=-1, verbose=0)
8 grid_svm_rbf.fit(X_train_flat, y_train)
9 best_svm_rbf = grid_svm_rbf.best_estimator_
10 print("SVM-RBF best params:", grid_svm_rbf.best_params_)
11
12 res_svm_rbf = evaluate_model(best_svm_rbf, X_train_flat, y_train, X_test_flat, y_test, "SVM-RBF")
13 results.append(res_svm_rbf)
14
...

```

✓ 3.3 Random Forest

```

1 rf = RandomForestClassifier(
2     n_estimators=300, max_depth=None, n_jobs=-1, random_state=42, class_weight=None
3 )
4 rf.fit(X_train_flat, y_train)
5 res_rf = evaluate_model(rf, X_train_flat, y_train, X_test_flat, y_test, "RandomForest")
6 results.append(res_rf)
7
...

```

✓ 3.4 Compare Classical Models

```

1 import pandas as pd
2 df_results = pd.DataFrame(results).sort_values(by="acc", ascending=False)
3 df_results
4
...

```

✓ add confusion matrices:

```

1 best_name = df_results.iloc[0]["model"]
2 best_clf = {
3     "LogisticRegression": best_lr,
4     "SVM-Linear": best_svm_lin,
5     "SVM-RBF": best_svm_rbf,
6     "RandomForest": rf
7 }[best_name]
8
9 y_pred_best = best_clf.predict(X_test_flat)
10 cm = confusion_matrix(y_test, y_pred_best)
11 plt.figure(figsize=(7,6))
12 sns.heatmap(cm, annot=False, cmap='Blues', xticklabels=class_names, yticklabels=class_names)
13 plt.xlabel("Predicted"); plt.ylabel("True"); plt.title(f"Confusion Matrix - {best_name}")
14 plt.tight_layout(); plt.show()
15

```

```
16 print(classification_report(y_test, y_pred_best, target_names=class_names))
...
```

✓ 4. CNN Model (Task 1, Deep Learning)

A compact CNN often outperforms classical models on Fashion-MNIST.

```
1 def make_cnn():
2     inputs = keras.Input(shape=(28,28,1))
3     x = layers.Conv2D(32, (3,3), activation='relu')(inputs)
4     x = layers.MaxPooling2D()(x)
5     x = layers.Conv2D(64, (3,3), activation='relu')(x)
6     x = layers.MaxPooling2D()(x)
7     x = layers.Flatten()(x)
8     x = layers.Dropout(0.3)(x)
9     x = layers.Dense(128, activation='relu')(x)
10    outputs = layers.Dense(num_classes, activation='softmax')(x)
11    model = keras.Model(inputs, outputs)
12    model.compile(optimizer='adam', loss='sparse_categorical_crossentropy', metrics=['accuracy'])
13    return model
14
15 cnn = make_cnn()
16 history = cnn.fit(
17     X_train_cnn, y_train,
18     validation_split=0.1,
19     epochs=8,
20     batch_size=128,
21     verbose=1
22 )
23
24 test_loss, test_acc = cnn.evaluate(X_test_cnn, y_test, verbose=0)
25 print(f"CNN - Accuracy: {test_acc:.4f}")
26
27 # make predictions on test data, verbose=0
```

```

27 y_proba_cnn = cnn.predict(x_test_cnn, verbose=0)
28 y_pred_cnn = np.argmax(y_proba_cnn, axis=1)
29
30 prec_macro_cnn = precision_score(y_test, y_pred_cnn, average='macro', zero_division=0)
31 y_true_bin = keras.utils.to_categorical(y_test, num_classes=num_classes)
32 auc_ovr_cnn = roc_auc_score(y_true_bin, y_proba_cnn, average='macro', multi_class='ovr')
33
34 print(f"CNN - Macro-Precision: {prec_macro_cnn:.4f} | OvR AUC: {auc_ovr_cnn:.4f}")
35

```

Add a markdown interpretation comparing CNN vs the best classical model:

Accuracy and macro-precision differences

CNN's ability to capture spatial structure vs flattened features

Optional: confusion matrix for CNN.

```

1 cm_cnn = confusion_matrix(y_test, y_pred_cnn)
2 plt.figure(figsize=(7,6))
3 sns.heatmap(cm_cnn, annot=False, cmap='Greens', xticklabels=class_names, yticklabels=class_names)
4 plt.xlabel("Predicted"); plt.ylabel("True"); plt.title("Confusion Matrix - CNN")
5 plt.tight_layout(); plt.show()
6
7 print(classification_report(y_test, y_pred_cnn, target_names=class_names))
8

```

5. Pick Best Model

```

1 # Consolidate with CNN metrics
2 row_cnn = {"model": "CNN", "acc": float(test_acc), "prec_macro": float(prec_macro_cnn), "auc_ovr": float(auc_ovr_cnn)}
3 df_all = pd.concat([df_results, pd.DataFrame([row_cnn])]).sort_values(by="acc", ascending=False)
4 df_all
5

```

Add markdown:

State the best model (likely CNN by accuracy and AUC).

Briefly justify selection and note any trade-offs (training time vs performance).

If the classical model somehow wins (rare), proceed with that instead for Task 2.

6. Task 2: Predict On Your Own 10 Images

Instructions:

Take 10 square photos (one per class if possible).

Save as PNG/JPG in a folder.

The pipeline: load → convert to grayscale → resize to 28x28 → normalize → shape to (28,28,1) → batch predict.

```

1 # Path with 10 images
2 user_img_dir = "YOUR_10_IMAGES_FOLDER" # e.g., "my_fashion_photos"
3 paths = sorted(glob.glob(os.path.join(user_img_dir, "*.jpg")))
4 assert len(paths) >= 10, f"Found {len(paths)} images, need at least 10."
5
6 def load_and_prepare_image(p, target=(28,28)):
7     img = Image.open(p).convert("L") # grayscale
8     img = img.resize(target) # ensure (28,28)
9     arr = np.array(img, dtype="float32")/255.0
10    return arr
11
12 user_imgs = [load_and_prepare_image(p) for p in paths[:10]]
13 user_stack_cnn = np.stack(user_imgs)[..., np.newaxis] # (N,28,28,1)
14

```

Visualize side-by-side:

```

1 plt.figure(figsize=(10,3))
2 for i, (p, arr) in enumerate(zip(paths[:10], user_imgs)):
3     plt.subplot(2,5,i+1)
4     plt.imshow(arr, cmap='binary')
5     plt.title(os.path.basename(p)[:12], fontsize=8)

```

```

6 plt.axis('off')
7 plt.tight_layout()
8 plt.show()
9

```

Predict with the best model (choose CNN if it won):

```

1 use_cnn = df_all.iloc[0]["model"] == "CNN"
2
3 if use_cnn:
4     probs = cnn.predict(user_stack_cnn, verbose=0)
5     preds = np.argmax(probs, axis=1)
6 else:
7     # If a classical model won:
8     user_stack_flat = user_stack_cnn.reshape(len(user_imgs), -1)
9     best_name = df_all.iloc[0]["model"]
10    best_clf = {
11        "LogisticRegression": best_lr,
12        "SVM-Linear": best_svm_lin,
13        "SVM-RBF": best_svm_rbf,
14        "RandomForest": rf
15    }[best_name]
16    preds = best_clf.predict(user_stack_flat)
17
18 pred_labels = [class_names[i] for i in preds]
19 pred_labels
20

```

If you have ground-truth labels for your images (optional), compute accuracy:

```

1 # Optional: set your true labels if known (list of 10 strings matching class_names)
2 # Example: y_true_names = ['Sneaker', 'Bag', ...]
3 y_true_names = None # replace with actual list
4
5 if y_true_names is not None:
6     true_idx = [class_names.index(n) for n in y_true_names]
7     pred_acc = accuracy_score(true_idx, preds)
8     print("User images accuracy:", pred_acc)
9

```

7. Short Discussion and Conclusions

Use markdown to:

Summarize which model outperformed and by how much.

Interpret metric differences (accuracy vs macro-precision vs AUC).

Mention practical considerations (speed, simplicity, deployment).

Note improvements: data augmentation, deeper CNN (BatchNorm, more epochs), regularization, or transfer learning variants.