MLPC_2025_Task3_Experiments

Team Laborer

Ali Doğukan Seven, Karel Štícha, Yisong Tang, Dmitrii Troitskii

2025S: 344.090

Model selection – Train 6 features on “speech” class

Linear SVM

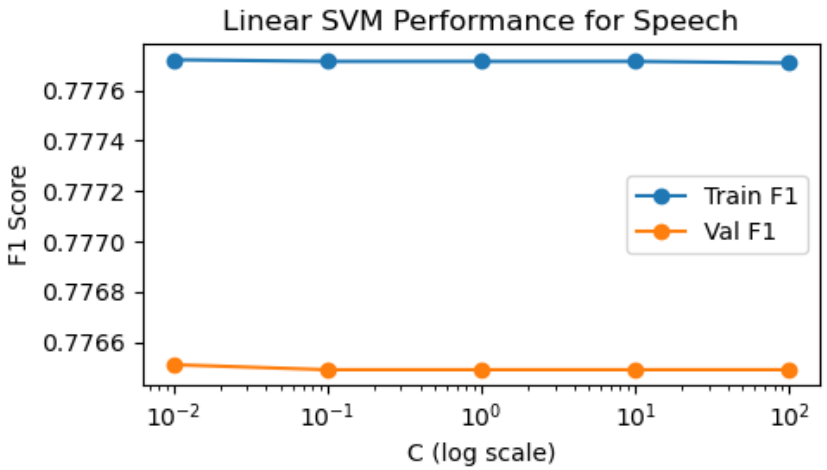
- Hyperparameter:
  do not have much impact on performance, we select C=1
- Performance:
  F1 remain almost constant (Train ≈ 0.7787, Val ≈ 0.7747)
- Over- vs. Under-fitting:
  over-fitting maybe
- Training time: all around 45s

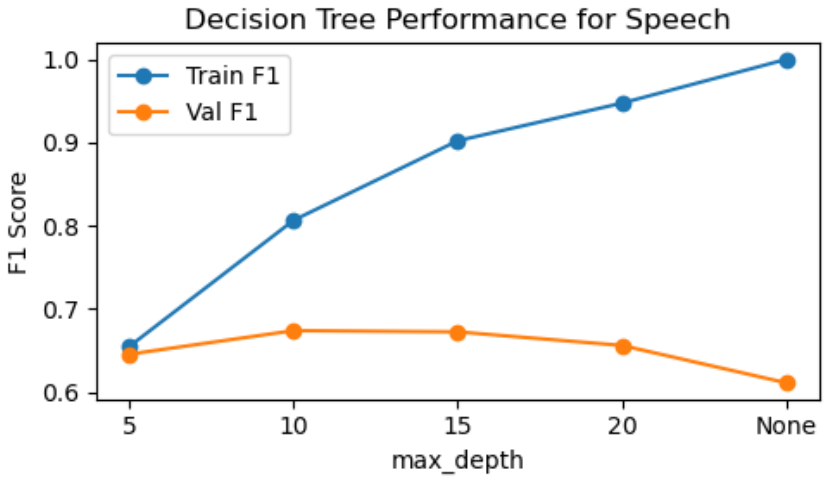
Decision Tree

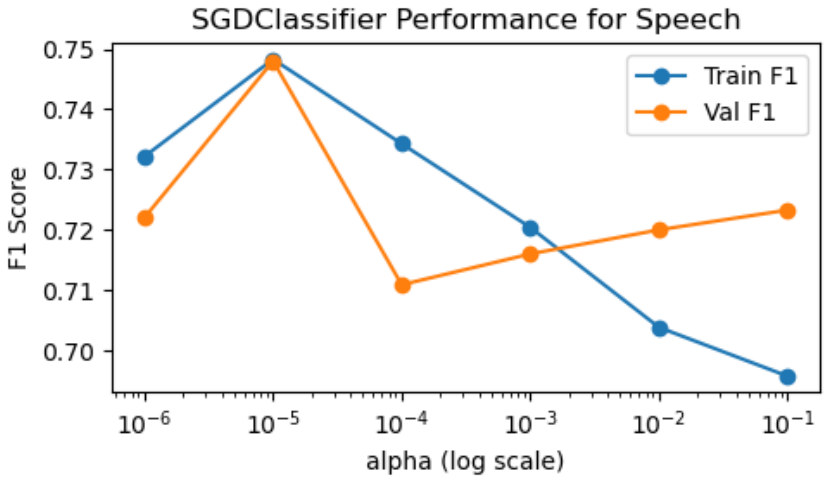
- Hyperparameter:
  Best max_depth=10
- Performance:
  best F1 ≈ 0.674 (Train ≈ 0.807)
- Over- vs. Under-fitting:
  when Depth > 10: over-fitting.
  When Depth <10: under-fitting
- Training time:
  increase from 223s to 2279s as the max_depth increase

Logistic Regression (SGD)

- Hyperparameter:
  Best α=1e-5
- Performance:
  Train F1 ≈ 0.7483 and Val F1 ≈ 0.7480
- Over- vs. Under-fitting:
  α >1e-5: under-fitting;
  α <1e-5: over-fitting.
- Training time： α=1e-4, almost 30 times to α=1e-5. (22.8s)
- α=1e-2,α=1e-1. training time ok.







Comparison of the Three Classifiers

Model	Best Hyperparameter	Validation F1	Training Speed
Linear SVM	C=1	0.775	~44 second per class
Logistic Regression(SGD)	$\alpha = 1e^{-5}$	0.748	5-90 seconds per class
Decision Tree	max_depth = 10	0.674	438 seconds per class

- Linear SVM achieves the highest Macro-F1 on the validation set with no over-fitting, making it the best choice.
- SGDClassifier is a close second, with very fast training—good for situations requiring frequent retraining.
- Decision Tree performs worst, is slow to train, and is sensitive to depth (easily under- or over-fits).

Final Model Performance

- Using the selected Linear SVM (C=1) trained on all 58 classes with the six features (mfcc, embeddings, contrast, flatness, bandwidth, melspectrogram), we evaluated on the held-out test set:
- Overall Macro-F1: 0.5381
- Per-class metrics: Precision, Recall, and F1 for each of the 58 classes were computed (see previous table).
- This final Macro-F1 reflects the average performance across all classes, balancing the model's ability to detect both frequent and rare sound events.

Class	Precision	Recall	F1 Score	Class	Precision	Recall	F1 Score
Fire	0.87	0.80	0.83	Horse Neigh	0.45	0.69	0.54
Stream/River	0.84	0.74	0.79	Hammer	0.69	0.44	0.54
Siren	0.80	0.77	0.78	Cow Moo	0.68	0.44	0.54
Train	0.88	0.69	0.77	Machine	0.53	0.51	0.52
Bird Chirp	0.85	0.71	0.77	Lawn Mower	0.57	0.46	0.51
Dog Bark	0.84	0.70	0.77	Beep/Bleep	0.60	0.44	0.50
Rain	0.80	0.73	0.76	Shout	0.64	0.42	0.50
Bell	0.90	0.66	0.76	Power Saw	0.65	0.39	0.49
Waves	0.78	0.73	0.76	Saxophone	0.48	0.46	0.47
Speech	0.81	0.70	0.75	Singing	0.85	0.32	0.47
Clapping	0.83	0.69	0.75	Drip	0.70	0.34	0.46
Guitar	0.85	0.66	0.74	Horn Honk	0.73	0.33	0.45
Rooster Crow	0.67	0.76	0.71	Violin	0.51	0.40	0.45
Snoring	0.60	0.82	0.69	Machine	0.47	0.43	0.45
Piano	0.70	0.69	0.69	Alarm	0.39	0.51	0.44
Insect Buzz	0.74	0.65	0.69	Car	0.61	0.33	0.43
Footsteps	0.84	0.58	0.68	Jackhammer	0.41	0.42	0.41
Thunder	0.68	0.63	0.65	Motorcycle	0.90	0.26	0.40
Laughter	0.76	0.56	0.64	Helicopter	0.50	0.33	0.39
Cat Meow	0.83	0.52	0.64	Cowbell	0.44	0.35	0.39
Bicycle	0.65	0.62	0.63	Wind	0.66	0.27	0.38
Drums	0.67	0.58	0.62	Power Drill	0.57	0.28	0.38
Vacuum Cleaner	0.60	0.64	0.62	Trumpet	0.39	0.24	0.30
Crying	0.68	0.54	0.60	Ship/Boat	0.37	0.22	0.28
Chainsaw	0.57	0.60	0.59	Hiccup	0.52	0.15	0.23
Cough	0.89	0.43	0.58	Sneeze	0.13	0.25	0.17
Airplane	0.81	0.45	0.58	Bus	0.90	0.04	0.07
Pig Oink	0.61	0.54	0.57	Truck	0.45	0.02	0.05
Sheep/Goat Bleat	0.63	0.50	0.56	Doorbell	0.02	0.17	0.03

Team Laborer

Thank You