**Building a Linear Classifier**

* How do we train a model?
* How do we use solution to create playlists?
* Idea: Train model on each example separately using unique intercept term, but same weights
* Problem: Which playlists is the “correct” one?

Original Data Good “Correct” Assignment Bad “Correct” Assignment

* Loss Function? Count # of songs assigned to wrong playlist
* Test possible assignments, choose one with best validation error
* Only want to capture relative position
* General Training Algorithm:

**Given assignment, generate classifiers**

p1,1 = 1, p1,2 = -1 y = **w** \* phi(x) + b1

p2,1 = -1, p2,2 = 1 y = **w** \* phi(x) + b2

p3,1 = -1, p3,2 = 1 y = **w** \* phi(x) + b3

**Given data, generate scores using weight** w

s1, s2, s3, s4, s5 ==> yi = **w** \* phi(si)

**Sort scores of songs:**

[y3, y1, y5, y4, y2]

**Split according to size of desired playlists:**

[s3, s1]

[s5, s4, s2]

* Try many random assignments, use one with best validation error

Data Set:

150 Playlist Pairs

80/20/20/30 Train/Validate 1/Validate 2/Test

Train 300 Random Assignments:

Average Train Error: 0.302

Average Validation Error: 0.347

Best Validation Error: 0.255

Test Error: 0.278

Random Assignment Expected Error: 0.44

**Better than random!**

* Check that better training error means better validation error
  + Plot Train Error vs. Validation Error (Good)
* Check if Validation Error is consistent
  + Plot Validation Error 1 vs. Validation Error 2 (Okay)

****