Cse 158 HW1

1. What is the variance of the ‘review/taste’ scores in the data? What is the Mean Squared Error (MSE) obtained when predicting the ‘review/taste’ score using the mean value of ‘review/taste’ (1 mark)?

Variance = 0.5135699375

MSE = 0.5135699375

2. How many reviews are there for each style of beer in the dataset (‘beer/style’)? What is the average value of ‘review/taste’ for reviews from each style? (1 mark)

Tripel : 257

avg review for Tripel : 3.78404669261

American Black Ale : 138

avg review for American Black Ale : 3.8731884058

Fruit / Vegetable Beer : 1355

avg review for Fruit / Vegetable Beer : 3.60774907749

English Brown Ale : 495

avg review for English Brown Ale : 3.72828282828

American Pale Ale (APA) : 2288

avg review for American Pale Ale (APA) : 3.64969405594

American Amber / Red Lager : 42

avg review for American Amber / Red Lager : 3.69047619048

Chile Beer : 11

avg review for Chile Beer : 3.95454545455

Cream Ale : 69

avg review for Cream Ale : 3.02898550725

Dubbel : 165

avg review for Dubbel : 3.73636363636

English Dark Mild Ale : 21

avg review for English Dark Mild Ale : 3.78571428571

Belgian Strong Pale Ale : 632

avg review for Belgian Strong Pale Ale : 4.05617088608

Weizenbock : 13

avg review for Weizenbock : 3.38461538462

Milk / Sweet Stout : 69

avg review for Milk / Sweet Stout : 3.78260869565

Dunkelweizen : 61

avg review for Dunkelweizen : 3.49180327869

Belgian Strong Dark Ale : 146

avg review for Belgian Strong Dark Ale : 3.69520547945

Foreign / Export Stout : 55

avg review for Foreign / Export Stout : 3.25454545455

Low Alcohol Beer : 7

avg review for Low Alcohol Beer : 2.71428571429

Euro Pale Lager : 701

avg review for Euro Pale Lager : 2.96291012839

Schwarzbier : 53

avg review for Schwarzbier : 3.62264150943

Munich Dunkel Lager : 141

avg review for Munich Dunkel Lager : 3.78014184397

Maibock / Helles Bock : 225

avg review for Maibock / Helles Bock : 3.74666666667

Belgian Pale Ale : 144

avg review for Belgian Pale Ale : 3.73958333333

Irish Dry Stout : 101

avg review for Irish Dry Stout : 3.62376237624

Märzen / Oktoberfest : 557

avg review for Märzen / Oktoberfest : 3.5933572711

California Common / Steam Beer : 11

avg review for California Common / Steam Beer : 3.31818181818

American Porter : 2230

avg review for American Porter : 4.08183856502

Vienna Lager : 33

avg review for Vienna Lager : 3.5303030303

Oatmeal Stout : 102

avg review for Oatmeal Stout : 3.77450980392

Bière de Garde : 7

avg review for Bière de Garde : 3.92857142857

Doppelbock : 873

avg review for Doppelbock : 3.98281786942

American Adjunct Lager : 242

avg review for American Adjunct Lager : 2.94834710744

Wheatwine : 455

avg review for Wheatwine : 4.18681318681

Scottish Ale : 78

avg review for Scottish Ale : 3.76282051282

English Porter : 367

avg review for English Porter : 3.70708446866

Saison / Farmhouse Ale : 141

avg review for Saison / Farmhouse Ale : 3.70212765957

Altbier : 165

avg review for Altbier : 3.40303030303

English India Pale Ale (IPA) : 175

avg review for English India Pale Ale (IPA) : 3.47142857143

Dortmunder / Export Lager : 31

avg review for Dortmunder / Export Lager : 3.41935483871

English Bitter : 267

avg review for English Bitter : 3.53745318352

Smoked Beer : 61

avg review for Smoked Beer : 3.19672131148

Lambic - Fruit : 6

avg review for Lambic - Fruit : 3.75

Munich Helles Lager : 650

avg review for Munich Helles Lager : 3.95923076923

American Barleywine : 825

avg review for American Barleywine : 4.06424242424

English Pale Ale : 1324

avg review for English Pale Ale : 3.48376132931

Belgian Dark Ale : 175

avg review for Belgian Dark Ale : 3.34

Rauchbier : 1938

avg review for Rauchbier : 4.06785345717

Hefeweizen : 618

avg review for Hefeweizen : 3.63511326861

Euro Strong Lager : 329

avg review for Euro Strong Lager : 2.84802431611

Baltic Porter : 514

avg review for Baltic Porter : 4.21303501946

Pumpkin Ale : 560

avg review for Pumpkin Ale : 3.7875

Herbed / Spiced Beer : 73

avg review for Herbed / Spiced Beer : 3.44520547945

Bock : 148

avg review for Bock : 3.18918918919

Euro Dark Lager : 144

avg review for Euro Dark Lager : 3.70486111111

American IPA : 4113

avg review for American IPA : 4.00085096037

Lambic - Unblended : 10

avg review for Lambic - Unblended : 3.3

English Barleywine : 133

avg review for English Barleywine : 4.36090225564

American Dark Wheat Ale : 14

avg review for American Dark Wheat Ale : 3.67857142857

Keller Bier / Zwickel Bier : 23

avg review for Keller Bier / Zwickel Bier : 3.86956521739

American Strong Ale : 166

avg review for American Strong Ale : 3.56927710843

English Stout : 136

avg review for English Stout : 3.59926470588

American Pale Lager : 123

avg review for American Pale Lager : 3.21544715447

Belgian IPA : 128

avg review for Belgian IPA : 3.94921875

American Brown Ale : 314

avg review for American Brown Ale : 3.74363057325

Braggot : 26

avg review for Braggot : 3.80769230769

American Double / Imperial IPA : 3886

avg review for American Double / Imperial IPA : 4.03332475553

American Wild Ale : 98

avg review for American Wild Ale : 4.1887755102

English Pale Mild Ale : 21

avg review for English Pale Mild Ale : 3.59523809524

Quadrupel (Quad) : 119

avg review for Quadrupel (Quad) : 3.59663865546

Light Lager : 503

avg review for Light Lager : 2.39662027833

Russian Imperial Stout : 2695

avg review for Russian Imperial Stout : 4.30037105751

Flanders Oud Bruin : 13

avg review for Flanders Oud Bruin : 3.92307692308

Berliner Weissbier : 10

avg review for Berliner Weissbier : 3.55

Scottish Gruit / Ancient Herbed Ale : 65

avg review for Scottish Gruit / Ancient Herbed Ale : 3.90769230769

Extra Special / Strong Bitter (ESB) : 667

avg review for Extra Special / Strong Bitter (ESB) : 3.68515742129

American Double / Imperial Stout : 5964

avg review for American Double / Imperial Stout : 4.47996311201

Scotch Ale / Wee Heavy : 2776

avg review for Scotch Ale / Wee Heavy : 4.08339337176

Winter Warmer : 259

avg review for Winter Warmer : 3.62162162162

Old Ale : 1052

avg review for Old Ale : 4.09600760456

German Pilsener : 586

avg review for German Pilsener : 3.66723549488

Kölsch : 94

avg review for Kölsch : 3.69680851064

Black & Tan : 122

avg review for Black & Tan : 3.94262295082

Flanders Red Ale : 2

avg review for Flanders Red Ale : 3.25

American Malt Liquor : 90

avg review for American Malt Liquor : 2.25555555556

Czech Pilsener : 1501

avg review for Czech Pilsener : 3.60959360426

Kristalweizen : 7

avg review for Kristalweizen : 2.78571428571

American Amber / Red Ale : 665

avg review for American Amber / Red Ale : 3.51353383459

American Pale Wheat Ale : 154

avg review for American Pale Wheat Ale : 3.33441558442

Rye Beer : 1798

avg review for Rye Beer : 4.21357063404

English Strong Ale : 164

avg review for English Strong Ale : 3.75609756098

American Blonde Ale : 357

avg review for American Blonde Ale : 3.25490196078

Witbier : 162

avg review for Witbier : 3.52777777778

American Stout : 591

avg review for American Stout : 3.81979695431

Irish Red Ale : 83

avg review for Irish Red Ale : 2.98192771084

Eisbock : 8

avg review for Eisbock : 3.75

American Double / Imperial Pilsner : 14

avg review for American Double / Imperial Pilsner : 3.82142857143

3. Train a simple predictor with a single binary feature indicating whether a beer is an ‘American IPA’:

review/taste ~ θ0+θ1×[beer is an American IPA]

Report the values of θ0 And θ1. Briefly describe your interpretation of these values, i.e., what do θ0 And θ1 represent (1 mark)?

theta: [θ0: 3.91520474 θ1: 0.08564622]

Both thetas represent the weight on given features to determine the score of review/taste. On the other hand, it is the line which the interception of y axis and the slope that has the minimized average distance with all sample points on the graph.

4. Split the data into two equal fractions – the first half for training, the second half for testing (based on the order they appear in the file). Train the same model as above on the training set only. What is the model’s MSE on the training and on the test set (1 mark)?

MSE(train): [ 0.55810729]

MSE(test): 0.468410050967

5. Extend the model above so that it incorporates binary features for

Every style of beer. Report the values of θ that you obtain, and the model’s MSE on the training and on the test set (1 mark).

Tripel 's thetas: [ 3.91103454 -0.24525914]

Tripel 's MSE(train): [ 0.55790879]

Tripel 's MSE(test): 0.469460683188

-------------------------------------

American Black Ale 's thetas: [ 3.90948103 -0.10645073]

American Black Ale 's MSE(train): [ 0.55832552]

American Black Ale 's MSE(test): 0.469160344682

-------------------------------------

Fruit / Vegetable Beer 's thetas: [ 3.91633805 -0.19631605]

Fruit / Vegetable Beer 's MSE(train): [ 0.557005]

Fruit / Vegetable Beer 's MSE(test): 0.465726670135

-------------------------------------

English Brown Ale 's thetas: [ 3.91088392 -0.26985827]

English Brown Ale 's MSE(train): [ 0.55790378]

English Brown Ale 's MSE(test): 0.468985032804

-------------------------------------

American Pale Ale (APA) 's thetas: [ 3.92184427 -0.28763135]

American Pale Ale (APA) 's MSE(train): [ 0.55487835]

American Pale Ale (APA) 's MSE(test): 0.465505293826

-------------------------------------

American Amber / Red Lager 's thetas: [ 3.90928734 -0.24262068]

American Amber / Red Lager 's MSE(train): [ 0.55833418]

American Amber / Red Lager 's MSE(test): 0.469062276429

-------------------------------------

Chile Beer 's thetas: [ 3.90918367 0.04081633]

Chile Beer 's MSE(train): [ 0.55835469]

Chile Beer 's MSE(test): 0.469125742607

-------------------------------------

Cream Ale 's thetas: [ 3.91046181 -0.95591636]

Cream Ale 's MSE(train): [ 0.55715077]

Cream Ale 's MSE(test): 0.468137802871

-------------------------------------

Dubbel 's thetas: [ 3.90982093 -0.16514008]

Dubbel 's MSE(train): [ 0.55825321]

Dubbel 's MSE(test): 0.46899751257

-------------------------------------

English Dark Mild Ale 's thetas: [ 3.90919094 0.02830906]

English Dark Mild Ale 's MSE(train): [ 0.5583551]

English Dark Mild Ale 's MSE(test): 0.469132395037

-------------------------------------

Belgian Strong Pale Ale 's thetas: [ 3.90499139 0.17025133]

Belgian Strong Pale Ale 's MSE(train): [ 0.55765655]

Belgian Strong Pale Ale 's MSE(test): 0.469510458693

-------------------------------------

Weizenbock 's thetas: [ 3.90947293 -0.52485754]

Weizenbock 's MSE(train): [ 0.55821219]

Weizenbock 's MSE(test): 0.469110947756

-------------------------------------

Milk / Sweet Stout 's thetas: [ 3.90901997 0.40916185]

Milk / Sweet Stout 's MSE(train): [ 0.55828173]

Milk / Sweet Stout 's MSE(test): 0.469955782121

-------------------------------------

Dunkelweizen 's thetas: [ 3.90922006 -0.04558369]

Dunkelweizen 's MSE(train): [ 0.55835445]

Dunkelweizen 's MSE(test): 0.469037204007

-------------------------------------

Belgian Strong Dark Ale 's thetas: [ 3.91015484 -0.1768215 ]

Belgian Strong Dark Ale 's MSE(train): [ 0.55818744]

Belgian Strong Dark Ale 's MSE(test): 0.468983687786

-------------------------------------

Foreign / Export Stout 's thetas: [ 3.90925274 -0.65925274]

Foreign / Export Stout 's MSE(train): [ 0.55832059]

Foreign / Export Stout 's MSE(test): 0.468214172745

-------------------------------------

Low Alcohol Beer 's thetas: [ 3.9095019 -1.5095019]

Low Alcohol Beer 's MSE(train): [ 0.55789973]

Low Alcohol Beer 's MSE(test): 0.4691928366

-------------------------------------

Euro Pale Lager 's thetas: [ 3.9232481 -1.23662838]

Euro Pale Lager 's MSE(train): [ 0.54118043]

Euro Pale Lager 's MSE(test): 0.462242045466

-------------------------------------

Schwarzbier 's thetas: [ 3.90938576 -0.17861653]

Schwarzbier 's MSE(train): [ 0.55832221]

Schwarzbier 's MSE(test): 0.46899911279

-------------------------------------

Munich Dunkel Lager 's thetas: [ 3.90984788 -0.80984788]

Munich Dunkel Lager 's MSE(train): [ 0.5578311]

Munich Dunkel Lager 's MSE(test): 0.472130542065

-------------------------------------

Maibock / Helles Bock 's thetas: [ 3.90934737 -0.40934737]

Maibock / Helles Bock 's MSE(train): [ 0.55829506]

Maibock / Helles Bock 's MSE(test): 0.46948718767

-------------------------------------

Belgian Pale Ale 's thetas: [ 3.90994089 -0.13926419]

Belgian Pale Ale 's MSE(train): [ 0.55825273]

Belgian Pale Ale 's MSE(test): 0.469028577295

-------------------------------------

Irish Dry Stout 's thetas: [ 3.90932767 -0.07599434]

Irish Dry Stout 's MSE(train): [ 0.55834567]

Irish Dry Stout 's MSE(test): 0.468976158828

-------------------------------------

Märzen / Oktoberfest 's thetas: [ 3.91104689 -0.52468325]

Märzen / Oktoberfest 's MSE(train): [ 0.55738974]

Märzen / Oktoberfest 's MSE(test): 0.468707123563

-------------------------------------

California Common / Steam Beer 's thetas: [ 3.90923459 -0.1235203 ]

California Common / Steam Beer 's MSE(train): [ 0.55835109]

California Common / Steam Beer 's MSE(test): 0.469070054881

-------------------------------------

American Porter 's thetas: [ 3.90832612 0.02909066]

American Porter 's MSE(train): [ 0.5583307]

American Porter 's MSE(test): 0.468371918831

-------------------------------------

Vienna Lager 's thetas: [ 3.90963518 -0.37515242]

Vienna Lager 's MSE(train): [ 0.55819229]

Vienna Lager 's MSE(test): 0.469075935243

-------------------------------------

Oatmeal Stout 's thetas: [ 3.90951178 -0.1498964 ]

Oatmeal Stout 's MSE(train): [ 0.55830872]

Oatmeal Stout 's MSE(test): 0.469082222375

-------------------------------------

Bière de Garde 's thetas: [ 3.90919457 0.01937685]

Bière de Garde 's MSE(train): [ 0.55835525]

Bière de Garde 's MSE(test): 0.469125403242

-------------------------------------

Doppelbock 's thetas: [ 3.91253416 -0.71856864]

Doppelbock 's MSE(train): [ 0.55597065]

Doppelbock 's MSE(test): 0.492915815593

-------------------------------------

American Adjunct Lager 's thetas: [ 3.91510117 -1.04630684]

American Adjunct Lager 's MSE(train): [ 0.55221575]

American Adjunct Lager 's MSE(test): 0.466040541291

-------------------------------------

Wheatwine 's thetas: [ 3.90929097 -0.28429097]

Wheatwine 's MSE(train): [ 0.55832951]

Wheatwine 's MSE(test): 0.47348904401

-------------------------------------

Scottish Ale 's thetas: [ 3.9092603 -0.0342603]

Scottish Ale 's MSE(train): [ 0.5583533]

Scottish Ale 's MSE(test): 0.469096397541

-------------------------------------

English Porter 's thetas: [ 3.91186482 -0.20689587]

English Porter 's MSE(train): [ 0.55781112]

English Porter 's MSE(test): 0.468928917765

-------------------------------------

Saison / Farmhouse Ale 's thetas: [ 3.90953432 -0.09286766]

Saison / Farmhouse Ale 's MSE(train): [ 0.55832442]

Saison / Farmhouse Ale 's MSE(test): 0.46897020129

-------------------------------------

Altbier 's thetas: [ 3.91172805 -0.4937593 ]

Altbier 's MSE(train): [ 0.5571135]

Altbier 's MSE(test): 0.468541362377

-------------------------------------

English India Pale Ale (IPA) 's thetas: [ 3.91207973 -0.43632215]

English India Pale Ale (IPA) 's MSE(train): [ 0.55710716]

English India Pale Ale (IPA) 's MSE(test): 0.468880496541

-------------------------------------

Dortmunder / Export Lager 's thetas: [ 3.90976947 -1.01691233]

Dortmunder / Export Lager 's MSE(train): [ 0.55777658]

Dortmunder / Export Lager 's MSE(test): 0.469720319606

-------------------------------------

English Bitter 's thetas: [ 3.91135851 -0.32904144]

English Bitter 's MSE(train): [ 0.55764978]

English Bitter 's MSE(test): 0.468255721587

-------------------------------------

Smoked Beer 's thetas: [ 3.91095698 -0.7706061 ]

Smoked Beer 's MSE(train): [ 0.55700451]

Smoked Beer 's MSE(test): 0.469153463402

-------------------------------------

Lambic - Fruit 's thetas: [ 3.90923822 -0.15923822]

Lambic - Fruit 's MSE(train): [ 0.55834928]

Lambic - Fruit 's MSE(test): 0.469123126524

-------------------------------------

Munich Helles Lager 's thetas: [ 3.90930186 -0.50930186]

Munich Helles Lager 's MSE(train): [ 0.55830349]

Munich Helles Lager 's MSE(test): 0.477238091031

-------------------------------------

American Barleywine 's thetas: [ 3.90468544 0.15851681]

American Barleywine 's MSE(train): [ 0.55766011]

American Barleywine 's MSE(test): 0.469256702103

-------------------------------------

English Pale Ale 's thetas: [ 3.91219855 -0.62469855]

English Pale Ale 's MSE(train): [ 0.55649116]

English Pale Ale 's MSE(test): 0.463169463102

-------------------------------------

Belgian Dark Ale 's thetas: [ 3.91321598 -0.57700909]

Belgian Dark Ale 's MSE(train): [ 0.55605423]

Belgian Dark Ale 's MSE(test): 0.468948937389

-------------------------------------

Rauchbier 's thetas: [ 3.90891784 0.14108216]

Rauchbier 's MSE(train): [ 0.55831563]

Rauchbier 's MSE(test): 0.46724624971

-------------------------------------

Hefeweizen 's thetas: [ 3.91296866 -0.31197528]

Hefeweizen 's MSE(train): [ 0.55719383]

Hefeweizen 's MSE(test): 0.46823863264

-------------------------------------

Euro Strong Lager 's thetas: [ 3.91872378 -1.14468532]

Euro Strong Lager 's MSE(train): [ 0.54754433]

Euro Strong Lager 's MSE(test): 0.464605861842

-------------------------------------

Baltic Porter 's thetas: [ 3.90890348 0.35300128]

Baltic Porter 's MSE(train): [ 0.55825078]

Baltic Porter 's MSE(test): 0.467392743466

-------------------------------------

Pumpkin Ale 's thetas: [ 3.90969377 -0.03578073]

Pumpkin Ale 's MSE(train): [ 0.55833794]

Pumpkin Ale 's MSE(test): 0.468950060466

-------------------------------------

Herbed / Spiced Beer 's thetas: [ 3.91013779 -0.68954955]

Herbed / Spiced Beer 's MSE(train): [ 0.55770959]

Herbed / Spiced Beer 's MSE(test): 0.469239827116

-------------------------------------

Bock 's thetas: [ 3.91243029 -1.04879392]

Bock 's MSE(train): [ 0.55497789]

Bock 's MSE(test): 0.469885646615

-------------------------------------

Euro Dark Lager 's thetas: [ 3.90959195 -0.18843811]

Euro Dark Lager 's MSE(train): [ 0.55828166]

Euro Dark Lager 's MSE(test): 0.468938772564

-------------------------------------

American IPA 's thetas: [ 3.90435639 0.05606027]

American IPA 's MSE(train): [ 0.55810729]

American IPA 's MSE(test): 0.468410050967

-------------------------------------

Lambic - Unblended 's thetas: [ 3.90944378 -0.60944378]

Lambic - Unblended 's MSE(train): [ 0.55820685]

Lambic - Unblended 's MSE(test): 0.469112454241

-------------------------------------

English Barleywine 's thetas: [ 3.9075271 0.46469512]

English Barleywine 's MSE(train): [ 0.55758077]

English Barleywine 's MSE(test): 0.468899794575

-------------------------------------

American Dark Wheat Ale 's thetas: [ 3.90932922 -0.2307578 ]

American Dark Wheat Ale 's MSE(train): [ 0.55832556]

American Dark Wheat Ale 's MSE(test): 0.469118391187

-------------------------------------

Keller Bier / Zwickel Bier 's thetas: [ 3.90919462 0.00747205]

Keller Bier / Zwickel Bier 's MSE(train): [ 0.55835532]

Keller Bier / Zwickel Bier 's MSE(test): 0.469126037269

-------------------------------------

American Strong Ale 's thetas: [ 3.91172095 -0.48479787]

American Strong Ale 's MSE(train): [ 0.55713956]

American Strong Ale 's MSE(test): 0.469577931654

-------------------------------------

English Stout 's thetas: [ 3.91097184 -0.40638468]

English Stout 's MSE(train): [ 0.55763845]

English Stout 's MSE(test): 0.46927602262

-------------------------------------

American Pale Lager 's thetas: [ 3.91090895 -0.89007561]

American Pale Lager 's MSE(train): [ 0.55683719]

American Pale Lager 's MSE(test): 0.468366624568

-------------------------------------

Belgian IPA 's thetas: [ 3.90900506 0.04274932]

Belgian IPA 's MSE(train): [ 0.55834706]

Belgian IPA 's MSE(test): 0.469135420287

-------------------------------------

American Brown Ale 's thetas: [ 3.90950728 -0.03507806]

American Brown Ale 's MSE(train): [ 0.55834468]

American Brown Ale 's MSE(test): 0.468989243769

-------------------------------------

Braggot 's thetas: [ 3.9093202 -0.13659293]

Braggot 's MSE(train): [ 0.55833896]

Braggot 's MSE(test): 0.469125808726

-------------------------------------

American Double / Imperial IPA 's thetas: [ 3.90658014 0.03620593]

American Double / Imperial IPA 's MSE(train): [ 0.55826737]

American Double / Imperial IPA 's MSE(test): 0.468140757127

-------------------------------------

American Wild Ale 's thetas: [ 3.90809975 0.28067576]

American Wild Ale 's MSE(train): [ 0.55804776]

American Wild Ale 's MSE(test): 0.469183763544

-------------------------------------

English Pale Mild Ale 's thetas: [ 3.90922732 -0.07589399]

English Pale Mild Ale 's MSE(train): [ 0.55835329]

English Pale Mild Ale 's MSE(test): 0.4690905722

-------------------------------------

Quadrupel (Quad) 's thetas: [ 3.9091891 0.0908109]

Quadrupel (Quad) 's MSE(train): [ 0.55835437]

Quadrupel (Quad) 's MSE(test): 0.469436138491

-------------------------------------

Light Lager 's thetas: [ 3.93927406 -1.5663574 ]

Light Lager 's MSE(train): [ 0.51215308]

Light Lager 's MSE(test): 0.467696549614

-------------------------------------

Russian Imperial Stout 's thetas: [ 3.86192141 0.44053853]

Russian Imperial Stout 's MSE(train): [ 0.53976258]

Russian Imperial Stout 's MSE(test): 0.473933574168

-------------------------------------

Flanders Oud Bruin 's thetas: [ 3.90919278 0.01388414]

Flanders Oud Bruin 's MSE(train): [ 0.55835526]

Flanders Oud Bruin 's MSE(test): 0.469125496923

-------------------------------------

Berliner Weissbier 's thetas: [ 3.90934374 -0.35934374]

Berliner Weissbier 's MSE(train): [ 0.55830373]

Berliner Weissbier 's MSE(test): 0.469117637563

-------------------------------------

Scottish Gruit / Ancient Herbed Ale 's thetas: [ 3.90903078 0.08135383]

Scottish Gruit / Ancient Herbed Ale 's MSE(train): [ 0.55834162]

Scottish Gruit / Ancient Herbed Ale 's MSE(test): 0.469165522263

-------------------------------------

Extra Special / Strong Bitter (ESB) 's thetas: [ 3.91027345 -0.15879416]

Extra Special / Strong Bitter (ESB) 's MSE(train): [ 0.55818605]

Extra Special / Strong Bitter (ESB) 's MSE(test): 0.468005981639

-------------------------------------

American Double / Imperial Stout 's thetas: [ 3.81913453 0.62930035]

American Double / Imperial Stout 's MSE(train): [ 0.50978892]

American Double / Imperial Stout 's MSE(test): 0.434675871343

-------------------------------------

Scotch Ale / Wee Heavy 's thetas: [ 3.88714593 0.20511597]

Scotch Ale / Wee Heavy 's MSE(train): [ 0.5543181]

Scotch Ale / Wee Heavy 's MSE(test): 0.471018609789

-------------------------------------

Winter Warmer 's thetas: [ 3.91161759 -0.28780806]

Winter Warmer 's MSE(train): [ 0.5576654]

Winter Warmer 's MSE(test): 0.468829365728

-------------------------------------

Old Ale 's thetas: [ 3.90888524 0.30265322]

Old Ale 's MSE(train): [ 0.5582602]

Old Ale 's MSE(test): 0.468325148244

-------------------------------------

German Pilsener 's thetas: [ 3.91009369 -0.93092702]

German Pilsener 's MSE(train): [ 0.5575242]

German Pilsener 's MSE(test): 0.47962613799

-------------------------------------

Kölsch 's thetas: [ 3.90986436 -0.20254728]

Kölsch 's MSE(train): [ 0.55822124]

Kölsch 's MSE(test): 0.469055183664

-------------------------------------

Black & Tan 's thetas: [ 3.9090361 0.03358685]

Black & Tan 's MSE(train): [ 0.55834988]

Black & Tan 's MSE(test): 0.469133702646

-------------------------------------

Flanders Red Ale 's thetas: [ 3.90921637 -0.40921637]

Flanders Red Ale 's MSE(train): [ 0.55834866]

Flanders Red Ale 's MSE(test): 0.469101198848

-------------------------------------

American Malt Liquor 's thetas: [ 3.90972389 -1.30972389]

American Malt Liquor 's MSE(train): [ 0.55766948]

American Malt Liquor 's MSE(test): 0.460360726845

-------------------------------------

Czech Pilsener 's thetas: [ 3.91195499 -0.59891151]

Czech Pilsener 's MSE(train): [ 0.55671295]

Czech Pilsener 's MSE(test): 0.470429931219

-------------------------------------

Kristalweizen 's thetas: [ 3.90943827 -0.9927716 ]

Kristalweizen 's MSE(train): [ 0.55811887]

Kristalweizen 's MSE(test): 0.469000512259

-------------------------------------

American Amber / Red Ale 's thetas: [ 3.91635476 -0.42587857]

American Amber / Red Ale 's MSE(train): [ 0.55535949]

American Amber / Red Ale 's MSE(test): 0.467547792117

-------------------------------------

American Pale Wheat Ale 's thetas: [ 3.91041424 -0.48184282]

American Pale Wheat Ale 's MSE(train): [ 0.55777176]

American Pale Wheat Ale 's MSE(test): 0.467659165296

-------------------------------------

Rye Beer 's thetas: [ 3.90884611 0.07434717]

Rye Beer 's MSE(train): [ 0.55832917]

Rye Beer 's MSE(test): 0.46630881364

-------------------------------------

English Strong Ale 's thetas: [ 3.9103111 -0.18155293]

English Strong Ale 's MSE(train): [ 0.55815487]

English Strong Ale 's MSE(test): 0.469118973612

-------------------------------------

American Blonde Ale 's thetas: [ 3.91539299 -0.76268363]

American Blonde Ale 's MSE(train): [ 0.55367042]

American Blonde Ale 's MSE(test): 0.467482995489

-------------------------------------

Witbier 's thetas: [ 3.91155286 -0.39477433]

Witbier 's MSE(train): [ 0.55743205]

Witbier 's MSE(test): 0.468983071787

-------------------------------------

American Stout 's thetas: [ 3.91127945 -0.10133793]

American Stout 's MSE(train): [ 0.55814896]

American Stout 's MSE(test): 0.469036076076

-------------------------------------

Irish Red Ale 's thetas: [ 3.9122191 -0.9434691]

Irish Red Ale 's MSE(train): [ 0.55551605]

Irish Red Ale 's MSE(test): 0.468952375366

-------------------------------------

Eisbock 's thetas: [ 3.90925096 -0.15925096]

Eisbock 's MSE(train): [ 0.55834725]

Eisbock 's MSE(test): 0.469122462469

-------------------------------------

American Double / Imperial Pilsner 's thetas: [ 3.90919637 0.09080363]

American Double / Imperial Pilsner 's MSE(train): [ 0.55835503]

American Double / Imperial Pilsner 's MSE(test): 0.469139182782

6. First, let’s train a predictor that estimates whether a beer is an ‘American IPA’ using two features: [‘beer/ABV’,‘review/taste’]. Train your predictor using an SVM classifier (see the code provided in class) – remember to train on the first half and test on the second half. Use a regularization constant of C= 1000 as in the code stub.

What is the accuracy (percentage of correct classifications) of the predictor on the train and test data?(1 mark)

train accuracy: 0.9226

test accuracy: 0.85632

7. Considering the ‘American IPA’ style, can you come up with a more accurate predictor (e.g. using features from the text, or otherwise)? Write down the feature vector you design, and report its train/test accuracy (1 mark).

[‘review/taste’, len(’beer/name’)]

|  |  |  |
| --- | --- | --- |
|  | [‘review/taste’, len(’beer/name’)] | ['beer/ABV','review/taste'] |
| train accuracy | 0.93264 | 0.9226 |
| test accuracy | 0.87492 | 0.85632 |

8. What effect does the regularization constant C have on the training/test performance? Report the train/test accuracy of your predictor from the previous question for

C ∈〈0.1,10,1000,100000〉.

Features: [ d['review/taste'], len(d['beer/name']) ]

|  |  |  |
| --- | --- | --- |
| C | Train accuracy | Test accuracy |
| 0.1 | 0.9294 | 0.92236 |
| 10 | 0.9322 | 0.88148 |
| 1000 | 0.93264 | 0.87492 |
| 100000 | 0.93272 | 0.874 |

The C affected the train and test accuracy, when C increase, test accuracy decrease, but train accuracy increase. Moreover, the higher of C, the less regularization because the model will try to increase the accuracy on training phrase but causes over-fitting and time consuming.