

データマイニング

回	日付け	内容	配付資料
#1	2011.10.03	ガイダンス, PDCA, データマイニングプロセス	配付資料#1
#2	2011.10.17	確率統計の復習, データに着せる既製服としての確率分布, 既製服のサイズの合わせ方としての学習基準	配付資料#2
#3	2011.10.24	線形学習機械, データとモデルの距離, 最小二乗法による学習	配付資料#3
#4	2011.10.31	線形学習機械の対極としてのk-最近接法, 学習と過学習, 統計的学習理論, 学習用データと検証用データ	
#5	2011.11.07		

演習課題 #0 2011.10.31

連絡事項

- このページが見られるようにしておいてください。
- Rのプログラムを走らせる環境を用意して下さい。
- レポート提出はワープロファイルでお願いしているので、ワープロとネット環境とメールアドレスも用意をお願いします。
- 今回の課題は、提出物はありません。

2011/11/07までに、下記のRのプログラムを動かせる環境を確保し、四角い枠の中をコピー&ペーストして、赤い文字が表示されない(エラーがでない)ことを確認しておいて下さい。

データファイル

```

x.1 <-
c(0.166680688260906, 0.690360480183903, -0.960737464041943, 0.153316713435719,
-0.83139709949313, -1.31291792000015, 0.354998775469648, 1.09532464417603,
0.400386940261498, 0.534457164420002, -0.214970828860256, 0.649688885350683,
0.270693725742763, -0.0970356673383812, -0.424063250183331, -0.0484267365569906,
0.914524377744286, 0.0545134303447268, -0.343362171657453, 1.62795245457545,
-0.710080258192599, -0.360308247076339, -0.316635995034297, -1.51896640831808,
-1.13723793881766, 2.05331402480514, -0.725777833640312, 1.47865446044515,
-0.316456080483368, 1.03016947449561, -1.06089790635716, -0.384964481829066,
0.189759682729100, 0.163071471425086, -0.288620545078447, -1.61379560709454,
0.233972451322242, 1.63459440000576, -0.107560016304604, 0.272599271179800,
2.85874531955234, -1.51744874470106, 0.245047724304523, -0.918285761964405,
1.31632412478431, -0.446719271851254, -1.76167737488009, -0.553148947434594,
0.000888062938795291, -0.0879856249093991, 1.88426733955180,
1.70181432563824, 1.82191525004163, 2.20463867315689, 2.90888860727657,
0.678279067206791, 2.20130956502179, 0.363250957269817, 1.24053875245629,
1.21972210174233, 0.265692467422776, 1.18505990790251, 2.98612037939113,
1.25276830938804, 1.93211551092542, 0.515716385482343, 0.991160344213232,
3.16467785962439, 1.58378730442831, 1.00525866635188, 1.14592228801568,
0.86739275499268, -1.13141770573282, 0.955422834963345, 2.09951460309889,
0.668306719923636, 1.42118523721167, 4.95868178952197, -0.42647051398233,
0.662686469132507, 2.50936264440344, 2.12616006480855, 3.27984026485583,
2.34145516926321, 3.34245771524733, 0.522262087288686, 0.900511977116232,
2.0046832747257, 4.18184036064689, 1.57945016419002, 2.28582563040758,
2.38305572075135, -0.0771693596667844, -0.430696939539103, 0.649357831276332,
2.59912130166155, 1.90398034068018, 4.02257056375588, 0.854114575781061,
2.24614141031886)
x.2 <-
c(-0.638105814357868, -0.735749207226363, 0.562003451633294,
-0.976277523232295, -0.633689897907844, -0.389764529607506, -0.408313673851167,
-1.24665598095712, -0.0473206079228157, -1.48609871281564, 0.156698075069995,
-1.35998111653076, 1.08853427032178, 0.568975635354217, 0.512839815789715,
-0.329649531671714, -0.325140978010364, 1.43124993209594, -1.45202379118742,
0.212757022650342, 1.23260329281572, 0.574264212359357, 0.175035168315614,
-1.06509013201713, -1.88583271910936, -0.827213788389483, -0.760679583613793,
0.846561646374465, 0.406509412548027, -0.572020121642617, -0.220112365715187,
1.11810061807253, -1.65755220957507, 0.187592425023049, 0.432607547497973,

```

```

1.01448803159590, -0.945864584796014, 0.281381974426747, 0.595120429090834,
-1.23749529149251, -0.206010596009017, 1.55368192304309, 0.56414901250267,
0.00449992567348487, -0.813169586850202, 0.0510458894915943,
-0.207768412925072, 0.505011250957093, 1.49492224046227, -0.681918102339157,
0.452152076545409, 2.26161140819322, 1.72057936292999, 2.11897296375225,
1.62841425164278, 0.461885535266931, 0.833424743941742, 1.95245907792897,
0.920624878877645, 1.34925600129103, 0.35120433604387, -1.12830573876754,
2.23774317944374, 0.176120829573867, 1.51689691539415, 0.0662186332757113,
1.31097141785189, 0.880119885722805, 0.757707040169653, 1.98895015674219,
1.95419111757019, 1.32921906632668, 0.00818902583933778, 1.21002814828982,
1.9465554930805, 1.69534863209990, 0.380252958275154, 2.99028468986211,
2.01408670602251, 1.72519225943921, 1.21730107117661, 1.58359324504399,
1.92504986594864, 2.17663698803132, 1.81690319813167, 0.877561459096729,
2.91213882209885, 0.300264821683087, 1.5379676839074, 0.166072552842766,
1.05259624522487, 0.604602117425379, 2.00667469704435, 2.59245439824437,
1.44472337802954, 0.791468866737462, 2.4435249987651, 0.166526271074167,
0.433030179321309, 0.0834871236892856)
Y <-
c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1)

```

最小二乗法による線形学習機械の学習。

```

lm(Y~X.1+X.2)
plot(X.1, X.2, col=(Y+1), pch=(Y+1), xlab="X1", ylab="X2", xlim=c(-2,5), ylim=c(-2,3), main="linear regression")
lines(c(-1.7,4), c((2496-1567*(-1.7))/2102, (2496-1567*(4))/2102))

```

k最近接法のプログラム

```

library(class)
Y.hat <- knn(cbind(X.1,X.2), cbind(X.1,X.2), Y, k=3, prob=TRUE)
X.test.1 <- c(-20:50)/10
X.test.2 <- c(-20:30)/10
X.test <- c(X.test.1[1], X.test.2[1])
for( i in 1:length(X.test.1) ) {
  for( j in 1:length(X.test.2) ) {
    X.test <- rbind(X.test, c(X.test.1[i], X.test.2[j]))
  }
}
X.test <- X.test[-1,]
# knn(k=3) vs lm
Y.hat <- knn(cbind(X.1,X.2), X.test, Y, k=3, prob=TRUE)
Y.hat.matrix <- matrix(0,nrow=71, ncol=51)
for( i in 1:length(X.test.1) ) {
  for( j in 1:length(X.test.2) ) {
    Y.hat.matrix[i,j] <- Y.hat[(i-1)*51+j]
  }
}
contour(X.test.1, X.test.2, Y.hat.matrix, xlab="X1", ylab="X2", lwd=1, nlevels=1, main="KNN with k=3")
points(X.1, X.2, col=(Y+1), pch=(Y+1))
lines(c(-2,4.2), c((2496-1567*(-2))/2102, (2496-1567*(4.2))/2102))

# knn(k=1) vs lm
Y.hat <- knn(cbind(X.1,X.2), X.test, Y, k=1, prob=TRUE)
Y.hat.matrix <- matrix(0,nrow=71, ncol=51)
for( i in 1:length(X.test.1) ) {
  for( j in 1:length(X.test.2) ) {
    Y.hat.matrix[i,j] <- Y.hat[(i-1)*51+j]
  }
}
contour(X.test.1, X.test.2, Y.hat.matrix, xlab="X1", ylab="X2", lwd=1, nlevels=1, main="KNN with k=1")
points(X.1, X.2, col=(Y+1), pch=(Y+1))
lines(c(-2,4.2), c((2496-1567*(-2))/2102, (2496-1567*(4.2))/2102))

# knn(k=7) vs lm
Y.hat <- knn(cbind(X.1,X.2), X.test, Y, k=7, prob=TRUE)
Y.hat.matrix <- matrix(0,nrow=71, ncol=51)
for( i in 1:length(X.test.1) ) {
  for( j in 1:length(X.test.2) ) {
    Y.hat.matrix[i,j] <- Y.hat[(i-1)*51+j]
  }
}

```

```
}  
}  
contour(X.test.1, X.test.2, Y.hat.matrix, xlab="X1", ylab="X2", lwd=1, nlevels=1, main="KNN with k=7")  
points(X.1, X.2, col=(Y+1), pch=(Y+1))  
lines(c(-2,4.2), c((2496-1567*(-2))/2102, (2496-1567*(4.2))/2102))  
  
# knn(k=15) vs lm  
Y.hat <- knn(cbind(X.1,X.2), X.test, Y, k=15, prob=TRUE)  
Y.hat.matrix <- matrix(0,nrow=71, ncol=51)  
for( i in 1:length(X.test.1) ) {  
  for( j in 1:length(X.test.2) ) {  
    Y.hat.matrix[i,j] <- Y.hat[(i-1)*51+j]  
  }  
}  
contour(X.test.1, X.test.2, Y.hat.matrix, xlab="X1", ylab="X2", lwd=1, nlevels=1, main="KNN with k=15")  
points(X.1, X.2, col=(Y+1), pch=(Y+1))  
lines(c(-2,4.2), c((2496-1567*(-2))/2102, (2496-1567*(4.2))/2102))
```

dmb/2011.txt · Last modified: 2011/10/31 10:02 by watalu

Except where otherwise noted, content on this wiki is licensed under the following license: [CC Attribution-Share Alike 3.0 Unported](http://creativecommons.org/licenses/by-sa/3.0/) [<http://creativecommons.org/licenses/by-sa/3.0/>]