

# Package ‘BinaryEMVS’

October 12, 2022

**Title** Variable Selection for Binary Data Using the EM Algorithm

**Version** 0.1

**Description** Implements variable selection for high dimensional datasets with a binary response variable using the EM algorithm. Both probit and logit models are supported. Also included is a useful function to generate high dimensional data with correlated variables.

**Depends** R (>= 3.1.3)

**License** GPL-3

**LazyData** true

**RoxygenNote** 5.0.1

**NeedsCompilation** no

**Author** John Snyder [aut, cre]

**Maintainer** John Snyder <jcs8v6@mail.missouri.edu>

**Repository** CRAN

**Date/Publication** 2016-01-13 08:49:37

## R topics documented:

BinomialEMVS . . . . .	1
data.sim . . . . .	3
<b>Index</b>	<b>4</b>

---

BinomialEMVS

*Variable Selection For Binary Data Using The EM Algorithm*

---

## Description

Conducts EMVS analysis

**Usage**

```
BinomialEMVS(y, x, type = "probit", epsilon = 5e-04, v0s = ifelse(type ==
  "probit", 0.025, 5), nu.1 = ifelse(type == "probit", 100, 1000),
  nu.gam = 1, lambda.var = 0.001, a = 1, b = ncol(x),
  beta.initial = NULL, sigma.initial = 1, theta.inital = 0.5, temp = 1,
  p = ncol(x), n = nrow(x), SDCD.length = 50)
```

**Arguments**

y	responses in 0-1 coding
x	X matrix
type	probit or logit model
epsilon	tuning parameter
v0s	tuning parameter, can be vector
nu.1	tuning parameter
nu.gam	tuning parameter
lambda.var	tuning parameter
a	tuning parameter
b	tuning parameter
beta.initial	starting values
sigma.initial	starting value
theta.inital	startng value
temp	not sure
p	not sure
n	not sure
SDCD.length	not sure

**Value**

probs is posterior probabilities

**Examples**

```
#Generate data
set.seed(1)
n=25;p=500;pr=10;cor=.6
X=data.sim(n,p,pr,cor)

#Randomly generate related beta coefficnets from U(-1,1)
beta.Vec=rep(0,times=p)
beta.Vec[1:pr]=runif(pr,-1,1)

y=scale(X%%beta.Vec+rnorm(n,0,sd=sqrt(3)),center=TRUE,scale=FALSE)
prob=1/(1+exp(-y))
y.bin=t(t(ifelse(rbinom(n,1,prob)>0,1,0)))
```

```
result.probit=BinomialEMVS(y=y.bin,x=X,type="probit")
result.logit=BinomialEMVS(y=y.bin,x=X,type="logit")

which(result.probit$posts>.5)
which(result.logit$posts>.5)
```

---

data.sim

*High Dimensional Correlated Data Generation*

---

### Description

Generates an high dimensional dataset with a subset of columns being related to the response, while controlling the maximum correlation between related and unrelated variables.

### Usage

```
data.sim(n = 100, p = 1000, pr = 3, cor = 0.6)
```

### Arguments

n	sample size
p	total number of variables
pr	the number of variables related to the response
cor	the maximum correlation between related and unrelated variables

### Value

Returns an  $n \times p$  matrix with the first  $pr$  columns having maximum correlation  $cor$  with the remaining  $p-pr$  columns

### Examples

```
data=data.sim(n=100,p=1000,pr=10,cor=.6)
max(abs(cor(data))[abs(cor(data))<1])
```

# Index

BinomialEMVS, 1

data.sim, 3