MATH 567: Mathematical Techniques in Data Science
Lab 5

Dominique Guillot

Departments of Mathematical Sciences
University of Delaware

March 8, 2017

Exercise 1: Multivariate normal distribution

1. Load the `mvtnorm` library.
2. Construct a mean vector $\mu = (0, 0)$ and a covariance matrix $\Sigma = X^T X$ for some random matrix $X \in \mathbb{R}^{2\times 2}$.
3. Construct vectors x, y and a matrix z as follows:

```
x = seq(-1,1,by=0.05)
y = seq(-1,1,by=0.05)
n = length(x)
z = matrix(0,nrow=n, ncol=n)
```

4. Compute $z[i, j] = f(x[i], y[j])$ where f is the density of $N(\mu, \Sigma)$. (Use the `dmvnorm` command).
5. Make a contour plot of the normal density: `contour(x,y,z)`.
6. Add the eigenvectors of Σ to the plot

```
e = eigen(Sigma)
arrows(0,0,e$vectors[1,1], e$vectors[2,1])
arrows(0,0,e$vectors[1,2], e$vectors[2,2])
```

7. Density in 3d: `persp(x,y,z,theta = 30, phi = 30)`.

Exercise 2: Titanic

1. Load the `titanic` training dataset (available on Sakai).
2. Split the data into a training set (2/3) and a test set (1/3).
3. Run `contrasts(data$Embarked)` to see how R handles categorical variables such as `Embarked`.
4. Train a logistic regression model to try to predict the fate of the passengers using some of the features:

```
model = glm(Survived ~ x1 + x2 + ... ,
family=binomial(link='logit'),data=train)
```

5. Predict values on the test set:

```
yhat = predict(model, test, type='response')
```

   (Note: returned values are of the form $P(Y = 1|X = x)$ because of the "response" option).

6. Compute the prediction accuracy of your model.