Exercises Pattern Recognition 2015
Linear Models for Regression and Classification

1 Linear Regression

An employer suspects that the performance of his employees depends on the temperature in the working environment according to the model

$$t = w_0 + w_1 x + \varepsilon$$

Here x denotes the temperature in degrees centigrade, and t denotes the performance of an employee (in some unit of measurement). The relationship is supposed to hold for $20 \leq x \leq 35$. To quantify this model, he collects the following 7 observations:

n	1	2	3	4	5	6	7
x_n	31	25	27	23	32	22	29
t_n	80	105	120	105	70	120	100

(a) Compute the least squares estimates of w_0 and w_1.

(b) Interpret the values of the estimates you have found under (a), that is, what do they mean?

(c) Use the fitted model to predict the productivity when the temperature in the working environment is 20 degrees centigrade.

(d) What percentage of the variation in performance is explained in this model by the variation in temperature?

2 Linear Models for Classification

It has often been claimed that the death penalty is applied in a racially discriminatory fashion. Data were provided by the Georgia Parole Board, the Georgia Supreme Court, and lawyers involved in the cases on the following variables:

variable	description
death	1 if got death penalty; 0 otherwise
blkdef	1 if black defendant; 0 otherwise
whtvict	1 if white victim; 0 otherwise
aggcirc	number of aggravating circumstances
fevict	1 if the victim is female; 0 otherwise
stranger	1 if victim is stranger; 0 otherwise
multvict	1 if 2 or more victims; 0 otherwise
multstab	1 if multiple stabs; 0 otherwise
yngvict	1 if victim 12 years or younger; 0 otherwise

We fitted a linear regression model to this data set, with `death` as the target variable. The results are summarized below:

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.18679    0.20034  -0.932 0.353609
blkdef      -0.08692    0.11024  -0.788 0.432482
whtvict      0.30522    0.12075   2.528 0.013202
aggcirc      0.06787    0.03714   1.827 0.070947
fevict       0.07903    0.10613   0.745 0.458409
stranger     0.35639    0.10146   3.512 0.000693
multvict     0.04994    0.13940   0.358 0.720987
multstab     0.28365    0.15177   1.869 0.064845
yngvict      0.05036    0.17730   0.284 0.777044

# summary of the fitted probabilities on the training data

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-0.1380  0.3489  0.5038  0.4900  0.6859  0.9320
```

(a) Which explanatory variables have a coefficient that is significantly different from zero at significance level $\alpha = 0.05$? And at $\alpha = 0.1$?

(b) According to this model, what is the probability that the defendant gets the death penalty when he or she is black, the victim is an asian man of 40 years old, the defendant and victim were good friends, the victim was strangled, and there were no aggravating circumstances.

(c) All else equal, according to this model, what is the difference in probability of the death penalty between a case where the victim and defendant knew each other and a case where victim and defendant were strangers?

(d) Interpreting the fitted coefficients and their p-values, would you say there is any evidence of racial discrimination in the application of the death penalty? Explain.

We also fitted a logistic regression model to the same data set. The results are summarized below:

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -3.5675     1.1243  -3.173 0.001508
blkdef       -0.5308     0.5439  -0.976 0.329059
whtvict       1.5563     0.6161   2.526 0.011528
aggcirc       0.3730     0.1963   1.900 0.057447
fevict        0.3707     0.5405   0.686 0.492829
stranger      1.7911     0.5386   3.325 0.000883
multvict      0.1999     0.7450   0.268 0.788490
multstab      1.4429     0.7938   1.818 0.069082
yngvict       0.1232     0.9526   0.129 0.897132

# summary of the fitted probabilities on the training data

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.03374 0.30220 0.48820 0.49000 0.71040 0.90180
```

(e) According to this model, what is the probability that the defendant gets the death penalty when the conditions are the same as under (b) of the previous question?

(f) Interpretation of the coefficients and the marginal effect of variables on the outcome is a bit more difficult than in the linear probability model. The fitted response function is given by

$$\hat{p}(t=1|\boldsymbol{x}) = (1 + e^{-\boldsymbol{w}_{\text{ML}}^\top \boldsymbol{x}})^{-1},$$

where $\boldsymbol{w}_{\text{ML}}$ are the maximum likelihood estimates of the coefficients $\boldsymbol{w} = [w_0, w_1, \ldots, w_{M-1}]^\top$. To assess the marginal effect of an increase in x_i on the fitted probability for class 1, determine:

$$\frac{\partial \hat{p}(t=1|\boldsymbol{x})}{\partial x_i},$$

where x_i is the i-th predictor variable, not the i-th observation of x.

3 Logistic Regression

In a study of commuting, for 21 persons their travel time to work by car and by public transport is determined. Also, each person in the study is asked whether he or she actually travels to work by car or public transport. Using these data, we estimate the model

$$p(t_n = 1 \mid x_n) = \frac{\exp(w_0 + w_1 x_n)}{1 + \exp(w_0 + w_1 x_n)},$$

where $t_n = 1$ means that person n travels to work by car, $t_n = 0$ that person n travels by public transport, and $x_n =$ (travel time by public transport $-$ travel time by car) for person n (in minutes). This produces the following maximum likelihood estimates

$$w_0 = -0.24 \qquad w_1 = 0.053$$

(a) We note that w_1 has a positive sign. Is this surprising? Explain.

(b) We also note that w_0 has a negative sign. Give a simple interpretation of this finding.

(c) According to this model, what is the probability that someone travels to work by car, if public transport takes 30 minutes longer?

(d) What is the marginal effect on the probability of choosing to travel by car, of an increase in x at $x = 5$? And at $x = 30$?

(e) Use the fitted model to give a simple classification rule for new cases.

4 Linear Regression and Logistic Regression

In the Google flu prediction example (see the lecture slides), first the logit (log-odds) transformation was performed on the target variable, and then a linear regression model was fitted to the transformed target. To obtain fitted probabilities from the model, the inverse transformation was performed. Couldn't we have applied the same strategy to the death penalty or travel data above? (with the advantage that we have a closed-form solution for the least-squares estimates in linear regression as opposed to the maximum likelihood estimates in logistic regression).

5 The Multinomial Logit Model

(a) Show that in the multinomial logit model we have:

$$\ln\left\{\frac{p(t=k|\mathbf{x})}{p(t=\ell|\mathbf{x})}\right\} = (\mathbf{w}_k - \mathbf{w}_\ell)^\top \mathbf{x},$$

(b) Show that the logistic regression model is a special case ($K = 2$) of the multinomial logit model. Use the identification constraint $\mathbf{w}_0 = \mathbf{0}$.

We have data on job categories, years of education, and whether or not someone belongs to a minority, of bank employees. The job categories are: 1=administrative job, 2=custodial job, 3=management job. We use years of education and minority (1 if person belongs to a minority; 0 otherwise) as explanatory variables for the job category. Note that category 1 (administrative job) is used as the reference category, that is, we put $\mathbf{w}_1 = \mathbf{0}$. The target variable ($t$) is `JOBCAT` and the two predictor variables are `EDUC` (x_1) and `MINORITY` (x_2).

```
> bankmen.multinom <- multinom(JOBCAT~EDUC+MINORITY,data=bankmen.dat)
# weights:  12 (6 variable)
initial  value 283.441970
iter  10 value 119.540307
iter  20 value 118.736632
final  value 118.736005
converged
> summary(bankmen.multinom)
Call:
multinom(formula = JOBCAT ~ EDUC + MINORITY, data = bankmen.dat)

Coefficients:
  (Intercept)       EDUC   MINORITY
2    4.760738 -0.5534006  0.4269483
3  -26.014242  1.6333793 -2.1090988

Std. Errors:
  (Intercept)       EDUC  MINORITY
2    1.172775 0.09904117 0.5027086
3    4.314470 0.27684953 0.7941959

Residual Deviance: 237.472
AIC: 249.472
```

Hence, we find that for example $w_{2,0} = 4.76$ (the constant term for category 2, rounded to two decimals), and $w_{3,2} = -2.11$ (the coefficient of `MINORITY` for category 3).

(c) According to the fitted model, what is the probability that an employee with 16 years of education who belongs to a minority (`MINORITY=1`) has a management job? Round the coefficient values to two decimal places in your calculations.

(d) We have found that $w_{3,1} = 1.63$. Interpret this coefficient value. Does the sign of the coefficient conform to common sense?