Statistical Learning Homework 3

* **Homework due: fifth, June 23:59**

1. Consider the following piecewise linear model:

   (a) The model is defined as:

   $$y = \begin{cases} \alpha_0 + \alpha_1 x + \epsilon, & \text{if } x \le x_0, \\ \beta_0 + \beta_1 x + \epsilon, & \text{if } x > x_0. \end{cases} \tag{1}$$

   If x_0 is known, the model can be rewritten as a linear model with a suitable choice of explanatory variables. Show that it can be expressed in the form:

   $$y = \alpha_0 + \alpha_1 Z_1 + \beta_1 Z_2 + \gamma_1 Z_3 + \epsilon. \tag{2}$$

   Identify Z_1, Z_2, Z_3, and γ_1.

   Note: Usually the change point x_0 is unknown, making the piecewise linear model nonlinear.

   (b) Suppose the linear model (2) is applied to the world population data in Table 1, using $x_0 = 1990$. Express the model in the matrix form $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ and identify the matrix $\mathbf{X}$ and vector $\boldsymbol{y}$.

   (c) According to the piecewise linear model in part (a), the conditional expectation $E(y|x)$ may be discontinuous at x_0. Find a condition involving α_0, α_1, β_0, β_1, and x_0 that ensures continuity at x_0.

Year	Population	Year	Population
1981	4.533	1991	5.367
1982	4.613	1992	5.450
1983	4.694	1993	5.531
1984	4.774	1994	5.611
1985	4.855	1995	5.691
1986	4.938	1996	5.769
1987	5.024	1997	5.847
1988	5.110	1998	5.925
1989	5.196	1999	6.003
1990	5.284	2000	6.080

Table 1: World Population Data (in billions)

(d) Rewrite the model from part (1), assuming continuity at x_0, in the following linear form:

$$y = \gamma_0 + \alpha_1 W_1 + \beta_1 W_2 + \epsilon \tag{3}$$

Identify γ_0, W_1, and W_2.

(e) If the linear model in part (d) is used for the world population data with $x_0 = 1990$, and the model is written as $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, identify the design matrix $\mathbf{X}$ and the associated parameter vector $\boldsymbol{\beta}$.

2. Do problem 5.8 in the textbook.

3. Do problem 6.11 in the textbook.