# Homework Sheet 5

**Group members:**
Dhvaniben Jasoliya
Leutrim Uka
Nicola Horst
Taqueer Rumaney
Yuvraj Dhepe

1. **6 Points**
   Calculate by hand two iterations of steepest gradient descent with line search for $\frac{1}{2}||X\beta - y||^2$ for $X = (2, 1; 1, 0)$ and $y = (1; 1)$ with initial iterate $\beta_0 = (1 \quad 1)^T$.

Rewrite $X, y, \beta_0$ using matrix notation:

$$X = \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix} \qquad y = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \beta_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

To solve the task, we use the algorithm as defined in the lecture

**Algorithm 1** Steepest Descent for Least-Squares

**for** $j = 1, \dots$ **do**
  Compute residual $\mathbf{r}_j = \mathbf{y} - \mathbf{X}\beta_j$
  Determine the SD direction $\mathbf{d}_j = \mathbf{X}^\top \mathbf{r}_j$
  Compute step size $\alpha_j = \frac{\mathbf{r}_j^\top \mathbf{X}\mathbf{d}_j}{\|\mathbf{X}\mathbf{d}_j\|^2}$
  Take the step $\beta_{j+1} = \beta_j + \alpha_j \mathbf{d}_j$
**end for**

Step 1 $(j = 0)$:

$$\rightarrow r_0 = y - X\beta_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \end{bmatrix}$$

$$\rightarrow d_0 = X^T r_0 = \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} -2 \\ 0 \end{bmatrix} = \begin{bmatrix} -4 \\ -2 \end{bmatrix}$$

$$\rightarrow \alpha_0 = \frac{r_0^T \cdot X \cdot d_0}{||Xd_0||^2} = \frac{20}{116} = 0,17$$

$$r_0^T X d_0 = \begin{bmatrix} -2 & 0 \end{bmatrix} \cdot \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} -4 \\ -2 \end{bmatrix} = \begin{bmatrix} -4 & -2 \end{bmatrix} \cdot \begin{bmatrix} -4 \\ -2 \end{bmatrix} = 16 + 4 = 20$$

$$||Xd_0||^2 = \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} -4 \\ -2 \end{bmatrix} = \begin{bmatrix} -10 \\ -4 \end{bmatrix} = (-10)^2 + (-4)^2 = 100 + 16 = 116$$

$$\rightarrow \beta_1 = \beta_0 + \alpha_0 d_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 0,17 \cdot \begin{bmatrix} -4 \\ -2 \end{bmatrix} = \begin{bmatrix} 0,31 \\ 0,66 \end{bmatrix}$$

Step 2 $(j = 1)$:

$$\rightarrow r_1 = y - X\beta_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0,31 \\ 0,66 \end{bmatrix} = \begin{bmatrix} -0,28 \\ 0,69 \end{bmatrix}$$

$$\rightarrow d_1 = X^T r_1 = \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} -0,27 \\ 0,68 \end{bmatrix} = \begin{bmatrix} 0,14 \\ 0,28 \end{bmatrix}$$

$$\rightarrow \alpha_1 = \frac{r_1^T X d_1}{||Xd_1||^2} = \frac{0,095}{0,02} = 4,9\overline{9}$$

$$r_1^T X d_1 = \begin{bmatrix} -0,28 & 0,69 \end{bmatrix} \cdot \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0,14 \\ 0,28 \end{bmatrix} = \begin{bmatrix} 0,14 & -0,28 \end{bmatrix} \cdot \begin{bmatrix} 0,14 \\ 0,28 \end{bmatrix} = 0,095$$

$$||Xd_1||^2 = 0,02$$

$$\rightarrow \beta_2 = \beta_1 + \alpha_1 d_1 = \begin{bmatrix} 0,31 \\ 0,66 \end{bmatrix} + 4,99 \begin{bmatrix} 0,14 \\ 0,28 \end{bmatrix} = \begin{bmatrix} 1 \\ -0,72 \end{bmatrix}$$

2. **6 points**
Show the solution to the so-called ridge regression is given by

$$\hat{\beta}_{\text{Ridge}} = \arg\min_{\beta} \|y - X\beta\|_2^2 + \lambda \|\beta\|_2^2 = (X^T X + \lambda I_p)^{-1} X^T y \tag{1}$$

To prove the equation above, we have to minimize the following expression:

$$\begin{aligned}
L(\beta) &= (y - X\beta)^T (y - X\beta) + \lambda \beta^T \beta \\
&= (y^T - \beta^T X^T)(y - X\beta) + \lambda \cdot \beta^T \beta \\
&= y^T y - y^T X\beta - \beta^T X^T y + \beta^T X^T X \beta + \lambda \cdot \beta^T \beta \\
&= \beta^T (X^T X + \lambda I)\beta - 2\beta^T X^T y + y^T y
\end{aligned}$$

To find the minimum, we have to derive the expression w.r.t. $\beta$:

$$\begin{aligned}
\frac{\partial}{\partial \beta} L(\beta) &= \frac{\partial}{\partial \beta}\left[\beta^T (X^T X + \lambda I)\beta - 2\beta^T X^T y + y^T y\right] \\
&= \frac{\partial}{\partial \beta}\left[\beta^T (X^T X + \lambda I)\beta\right] - 2 \cdot \frac{\partial}{\partial \beta} \beta^T X^T y + 0 \\
&= 2(X^T X + \lambda I)\beta - 2X^T y
\end{aligned}$$

Set $\frac{\partial}{\partial \beta} L(\beta) = 0$:

$$2(X^T X + \lambda I)\beta - 2X^T y = 0$$

$$\cancel{2}(X^T X + \lambda I)\beta = \cancel{2}X^T y$$

$$(X^T X + \lambda I)\beta = X^T y$$

$$\beta = \frac{X^T y}{X^T X + \lambda I}$$

$$\boxed{\beta = (X^T X + \lambda I)^{-1} X^T y}$$