Tutorial: Linear Models for Regression

- **Author: Yunqiu Xu**

Question 1 A *univariate linear regression model* is a linear equation $y = a + bx$. Learning such a model requires fitting it to a sample of training data so as to minimize the error function $E(a, b) = \sum_{i=1}^{n}(y_i - (a + bx_i))^2$. To find the best parameters a and b that minimize this error function we need to find the error *gradients* $\frac{\partial E(a,b)}{\partial a}$ and $\frac{\partial E(a,b)}{\partial b}$. So derive these expressions by taking partial derivatives, divide by n (the number of (x, y) data points in the training sample) set them to zero, and solve for a and b.

- Solution:

In order to find the parameters we take partial derivatives, set the partial derivatives to 0 and solve for a and b:

$$\frac{\partial}{\partial a}\sum_{i=1}^{n}(w_i - (a + bh_i))^2 = -2\sum_{i=1}^{n}(w_i - (a + bh_i)) = 0$$

$$\Rightarrow \hat{a} = \overline{w} - \hat{b}\overline{h}$$

$$\frac{\partial}{\partial b}\sum_{i=1}^{n}(w_i - (a + bh_i))^2 = -2\sum_{i=1}^{n}(w_i - (a + bh_i))h_i = 0$$

$$\Rightarrow \hat{b} = \frac{\sum_{i=1}^{n}(h_i - \overline{h})(w_i - \overline{w})}{\sum_{i=1}^{n}(h_i - \overline{h})^2}$$

So the solution found by linear regression is $w = \hat{a} + \hat{b}h = \overline{w} + \hat{b}(h - \overline{h})$.

Question 2 A linear regression model is represented by the linear equation $y = a + bx$. Show that the mean point $(\bar{x}, \bar{y})$ must line on the regression line.

- Solution: as we have derived $a = \overline{y} - b\overline{x}$, we can then get $\overline{y} = a + b\overline{x}$, which means the mean point must be on the regression line

---

Question 3 *Mean Square Error*, or MSE, of an estimator such as a regression model can be decomposed. Show that MSE = (variance) + (bias)^2.

- Some related equations are given:

$$E[X] = \sum_{i=1}^{N} x_i \, p(x_i) = \mu$$

$$\text{MSE} = E[f - y]^2$$

$$E[f(X)] = \sum_{i=1}^{N} f(x_i) \, p(x_i)$$

$$\begin{aligned}
E[fE[y]] &= fE[y] && \text{for } f \text{ a deterministic function, } y \text{ a random variable} \\
E[E[y]] &= E[y] \\
E[E[y]^2] &= E[y]^2 \\
E[yf] &= fE[y] && \text{again, since } f \text{ is deterministic, and } y \text{ a random variable} \\
E[yE[y]] &= E[y]^2 && \text{by the definition of expectation.}
\end{aligned}$$

- Solution:
- $Var = E[y - E[y]]^2 = E[y^2 - 2yE[y] + E[y]^2]$
- $Bias^2 = (E[y] - f)^2 = E[y]^2 - 2fE[y] + f^2$
-

$$\begin{aligned}
MSE &= E[f - y]^2 \\
&= E[f^2 - 2fy + y^2] \\
&= E[(y^2 - 2yE[y] + E[y]^2) + (E[y]^2 - 2fE[y] + f^2) \\
&\quad + 2(yE[y] - E[y]^2 + fE[y] - fy)] \\
&= Var \ + \ Bias^2 \ + \ 2E[yE[y] - E[y]^2 + fE[y] - fy] \\
&= Var \ + \ Bias^2
\end{aligned}$$