Tutorial: Numeric Prediction

Machine Learning

Numeric Prediction I

1. Show $\text{MSE} = (\text{variance}) + (\text{bias})^2$

2. To find a linear model $Y = a + bX$ that minimises mean-square error given points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, we first want to find the gradients $\frac{\partial Y}{\partial b}$ and $\frac{\partial Y}{\partial a}$. Show:

$$\nabla_b = \frac{\partial Y}{\partial b} = -\frac{2}{n}\sum_i^n x_i(y_i - (a + bx_i))$$

$$\nabla_a = \frac{\partial Y}{\partial a} = -\frac{2}{n}\sum_i^n (y_i - (a + bx_i))$$

3. We want to fit a linear model $Y = a + bX$. Derive expressions for a and b that do this.

Numeric Prediction II

4. We want to fit a linear model $Y = a + bX$ by minimising mean-square error. We know that a minimum occurs when $\nabla_a = 0$ and $\nabla_b = 0$. Show, by setting $\nabla_a = 0$ that the point $(\overline{X}, \overline{Y})$ lies on the regression line (that is, $\overline{Y} = a + b\overline{X}$)

5. We want to fit a linear model $Y = a + bX$ by finding a and b using gradient descent. Write the iterative update equations for a and b in terms of the gradients

6. What changes with gradient descent we want to fit a non-linear model $Y = a + bX + cX^2$?

7. What happens the the cost function being minimised has multiple local minima?

8. Is the cost function being minimised in least-squares regression convex?

Numeric Prediction III

9. We want to fit a linear model $Y = a + bX$. For the special case that the erorrs $e_i \sim_{\mathrm{i}.i.d.} N(0, \sigma^2)$ show that the least-square estimate for b is the same as the maximimum likelihood estimate for b

10. Derive equations for gradient descent when a regularisation term is added to the usual MSE cost function.