Tutorial 5 - Numeric Prediction

25th Feb, 2020

1. Show MSE $=$ (variance) $+$ (bias)^2

MSE of an estimator is defined as:

$$\operatorname{MSE}(\hat{\theta}) = \operatorname{E}_\theta\left[(\hat{\theta} - \theta)^2\right]$$

Ans.

$$\begin{aligned}
\operatorname{MSE}(\hat{\theta}) &= \operatorname{E}_\theta\left[(\hat{\theta} - \theta)^2\right] \\
&= \operatorname{E}_\theta\left[\left(\hat{\theta} - \operatorname{E}_\theta[\hat{\theta}] + \operatorname{E}_\theta[\hat{\theta}] - \theta\right)^2\right] \\
&= \operatorname{E}_\theta\left[\left(\hat{\theta} - \operatorname{E}_\theta[\hat{\theta}]\right)^2 + 2\left(\hat{\theta} - \operatorname{E}_\theta[\hat{\theta}]\right)\left(\operatorname{E}_\theta[\hat{\theta}] - \theta\right) + \left(\operatorname{E}_\theta[\hat{\theta}] - \theta\right)^2\right] \\
&= \operatorname{E}_\theta\left[\left(\hat{\theta} - \operatorname{E}_\theta[\hat{\theta}]\right)^2\right] + \operatorname{E}_\theta\left[2\left(\hat{\theta} - \operatorname{E}_\theta[\hat{\theta}]\right)\left(\operatorname{E}_\theta[\hat{\theta}] - \theta\right)\right] + \operatorname{E}_\theta\left[\left(\operatorname{E}_\theta[\hat{\theta}] - \theta\right)^2\right]
\end{aligned}$$

$$
\begin{aligned}
&= \mathrm{E}_\theta\left[\left(\hat{\theta} - \mathrm{E}_\theta[\hat{\theta}]\right)^2\right] + 2\left(\mathrm{E}_\theta[\hat{\theta}] - \theta\right)\mathrm{E}_\theta\left[\hat{\theta} - \mathrm{E}_\theta[\hat{\theta}]\right] + \left(\mathrm{E}_\theta[\hat{\theta}] - \theta\right)^2 \\
&= \mathrm{E}_\theta\left[\left(\hat{\theta} - \mathrm{E}_\theta[\hat{\theta}]\right)^2\right] + 2\left(\mathrm{E}_\theta[\hat{\theta}] - \theta\right)\left(\mathrm{E}_\theta[\hat{\theta}] - \mathrm{E}_\theta[\hat{\theta}]\right) + \left(\mathrm{E}_\theta[\hat{\theta}] - \theta\right)^2 \\
&= \mathrm{E}_\theta\left[\left(\hat{\theta} - \mathrm{E}_\theta[\hat{\theta}]\right)^2\right] + \left(\mathrm{E}_\theta[\hat{\theta}] - \theta\right)^2 \\
&= \mathrm{Var}_\theta(\hat{\theta}) + \mathrm{Bias}_\theta(\hat{\theta}, \theta)^2
\end{aligned}
$$

2. Find a linear model Y = a + bX that minimises mean-square error given points (x1, y1),(x2, y2), . . . ,(xn, yn).

Step 1: Find $\nabla_{\mathbf{a}}$ and $\nabla_{\mathbf{b}}$

Step 2: Show, by setting $\nabla_{\mathbf{a}} = 0$ that the point $(\overline{X}, \overline{Y})$ lies on the regression line (that is, Y = a + bX)

Step 3: Derive expressions for a and b.

3. 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.

Ans.

$$b_{k+1} = b_k - \eta \nabla_b$$

$$a_{k+1} = a_k - \eta \nabla_a$$

Here, η is the step-size.

4. What changes with gradient descent we want to fit a

non-linear model $Y = a + bX + cX^2$?

Ans. Algorithmically, nothing. We just have to find the extra gradient ∇c and use the corresponding update equation for c.

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

Ans. Gradient descent can get stuck in a local minimum that can be far away from the global minimum. **Random restarts** will not provably fix this, although it may help.

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

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

THANK YOU!