

2413, Machine Learning, Tutorial 7

Universität Bern

Simon Jenni (jenni@inf.unibe.ch)

Mutual Information

1. Consider the following set of 2-dimensional points, sampled from two classes:

x_1	x_2	y
1	0	1
0	0	1
0	1	1
1	1	1
0	0	0
1	0	0
0	1	0
0	0	0

Find the feature x_i with the highest mutual information

$$\text{MI}(x_i, y) = \sum_{x_i \in \{0,1\}} \sum_{y \in \{0,1\}} p(x_i, y) \log \frac{p(x_i, y)}{p(x_i)p(y)}. \quad (1)$$

Regularization

2. Consider the problem of linear regression with a Gaussian prior on the parameter vector θ of the form $p(\theta) = \mathcal{N}(0, \lambda I)$, where I is the identity matrix. Derive the cost function for the **MAP** estimate θ_{MAP} .

Hint: Remember the linear regression assumption $y^{(i)} = \theta^T x^{(i)} + \epsilon^{(i)}$ where $\epsilon^{(i)} \sim \mathcal{N}(0, \sigma)$.

3. What is the gradient of the cost derived in the previous question?