MLISP: Machine Learning in Signal Processing

Solutions to problem set 2

Prof. V. I. Morgenshtern

Solver: S. Lotter and V. I. Morgenshtern

Problem 1: Probabilistic interpretation of least-squares criterion

1. Since $\mathbf{x}^{(i)}$ are nonrandom, the randomness in $\mathrm{y}^{(i)}$ is due to the Gaussian noise only. Therefore,

$$p_{\mathrm{y}^{(i)}|\boldsymbol{\theta}}(y^{(i)}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left(y^{(i)} - \boldsymbol{\theta}^\mathsf{T}\mathbf{x}^{(i)}\right)^2}{2\sigma^2}\right).$$

2. The independence of $\mathrm{n}^{(i)}$ over i implies the independence of $\mathrm{y}^{(i)}$ over i. Therefore,

$$p_{\mathbf{y}|\boldsymbol{\theta}}(\mathbf{y}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left(y^{(i)} - \boldsymbol{\theta}^\mathsf{T}\mathbf{x}^{(i)}\right)^2}{2\sigma^2}\right) = \frac{1}{(2\pi)^{\frac{n}{2}}\,\sigma^n} \exp\left(-\sum_{i=1}^{n} \frac{\left(y^{(i)} - \boldsymbol{\theta}^\mathsf{T}\mathbf{x}^{(i)}\right)^2}{2\sigma^2}\right).$$

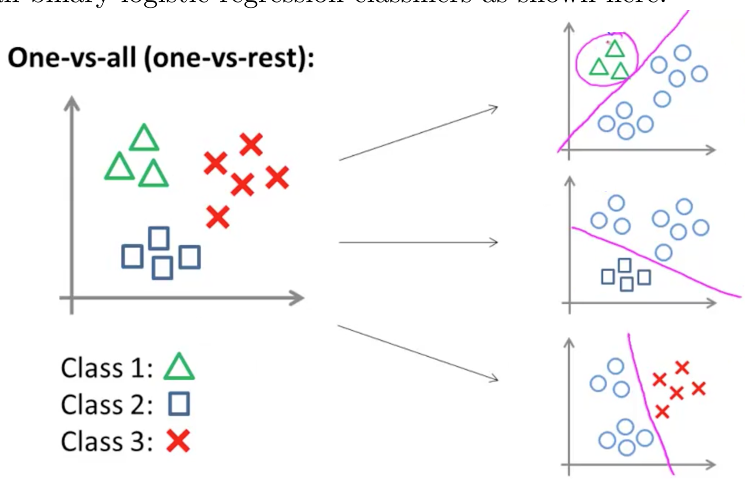
3. Using the monotonicity of the natural logarithm function

$$\begin{aligned}
\arg\max_{\boldsymbol{\theta}} p_{\mathbf{y}|\boldsymbol{\theta}}(\mathbf{y}) &= \arg\max_{\boldsymbol{\theta}} \ln\bigl(p_{\mathbf{y}|\boldsymbol{\theta}}\bigr) \\
&= \arg\max_{\boldsymbol{\theta}} \left\{\ln\left(\frac{1}{(2\pi)^{\frac{n}{2}}\,\sigma^n}\right) + \sum_{i=1}^{n} -\frac{\left(y^{(i)} - \boldsymbol{\theta}^\mathsf{T}\mathbf{x}^{(i)}\right)^2}{2\sigma^2}\right\} \\
&= \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \left(y^{(i)} - \boldsymbol{\theta}^\mathsf{T}\mathbf{x}^{(i)}\right)^2
\end{aligned}$$

which is the least-squares criterion.

Problem 7: Multi-class classification (exam practice)

Train three one-vs-all binary logistic regression classifiers as shown here:



Recall that

$$h_{\boldsymbol{\theta}}^{(i)}(\mathbf{x}) = P\left(y = i|\mathbf{x}, \boldsymbol{\theta}^{(i)}\right), \ i = 1, 2, 3.$$

At prediction time:

$$\hat{i} = \arg\max_{i} h_{\boldsymbol{\theta}}^{(i)}(\mathbf{x}).$$