

CSE 455/555 Spring 2013 Homework 7: Parametric Techniques

Jason J. Corso
Computer Science and Engineering
SUNY at Buffalo
jcorso@buffalo.edu

This assignment does not need to be submitted and will not be graded, but students are advised to work through the problems to ensure they understand the material.

You are both allowed and encouraged to work in groups on this and other homework assignments in this class. These are challenging topics, and working together will both make them easier to decipher and help you ensure that you truly understand them.

1. Gaussian Distribution

The univariate Gaussian have the following form:

$$\mathcal{N}(x|\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x - \mu)^2\right\}$$

Now, we are given N one dimensional independently and identically distributed (i.i.d.) samples $\mathbf{X} = \{x_1, x_2, \dots, x_N\}$ from a Gaussian distribution $\mathcal{N}(x|\mu, \sigma^2)$, and we want to determine the parameter from this set of samples.

- What is the joint probability distribution $P(\mathbf{X}|\mu, \sigma^2)$ of the samples?
- What is the maximum likelihood (ML) estimation of the parameters, if both μ and σ^2 are unknown?
- Show that the ML estimation of the variance is biased, *i.e.* show that

$$E(\sigma_{ML}^2) = \frac{N-1}{N}\sigma^2$$

Hint: you can use the fact that the expectation of random variable from a Gaussian distribution is μ , *i.e.* $\mathbb{E}(x) = \mu$, and $\text{var}(x) = \sigma^2 = \mathbb{E}(x^2) - \mathbb{E}(x)^2$

- Please write down the objective function of maximum a posteriori (MAP) estimation of the parameters, if we assume that only μ is the unknown parameter and follow a Gaussian distribution with mean μ_0 and variance σ_0^2 .
- Please write down the Bayesian formulation of this problem, if all other assumptions stay the same as in question D.

2. Naive Bayes

In naive Bayes, we assume that the presence of a particular feature of a class is unrelated to the presence of any other feature. For example, a fruit may be considered to be a watermelon if it is green, round, and more than 10 pounds. we will consider all these three features as independent to each other in naive Bayes.

Let our features $x_i, i \in [1, d]$ be binary valued and have d dimensions, *i.e.* $x_i \in \{0, 1\}$ and our input feature vector $x = [x_1 x_2 \dots x_d]^T$. For each training sample, our target value $y \in \{0, 1\}$ is also a binary-valued

variable. Then our model is parameterized by $\phi_{i|y=0} = p(x_i = 1|y = 0)$, $\phi_{i|y=1} = p(x_i = 1|y = 1)$, and $\phi_y = p(y = 1)$, and

$$\begin{aligned}
 p(y) &= (\phi_y)^y (1 - \phi_y)^{(1-y)} \\
 p(x|y = 0) &= \prod_{i=1}^d p(x_i|y = 0) \\
 &= \prod_{i=1}^d (\phi_{i|y=0})^{x_i} (1 - \phi_{i|y=0})^{(1-x_i)} \\
 p(x|y = 1) &= \prod_{i=1}^d p(x_i|y = 1) \\
 &= \prod_{i=1}^d (\phi_{i|y=1})^{x_i} (1 - \phi_{i|y=1})^{(1-x_i)}
 \end{aligned}$$

- Write down the joint log-likelihood function $l(\theta) = \log \prod_{n=1}^N p(x^{(n)}, y^{(n)}; \theta)$ in terms of the model parameters given above. $x^{(n)}$ means the n th data point, and θ represents all the parameters, *i.e.* $\{\phi_y, \phi_{i|y=0}, \phi_{i|y=1}, i = 1, \dots, d\}$.
- Estimate the parameters using maximum likelihood, *i.e.* find solutions for parameter ϕ_y , $\phi_{i|y=0}$ and $\phi_{i|y=1}$.
- When a new sample point x comes, we make the prediction based on the most likely class estimate generated by our model. Show that the hypothesis returned by naive Bayes is linear, *i.e.* if $p(y = 0|x)$ and $p(y = 1|x)$ are the class probabilities returned by our model, show that there exist some α so that

$$p(y = 1|x) \geq p(y = 0|x) \text{ if and only if } \alpha^{\mathbf{T}} \tilde{\mathbf{x}} \geq 0$$

where $\alpha = [\alpha_0 \ \alpha_1 \ \dots \ \alpha_d]^{\mathbf{T}}$ and $\tilde{\mathbf{x}} = [1 \ x_1 \ x_2 \ \dots \ x_d]^{\mathbf{T}}$.