Zoltan Onodi-Szucs

zo2131

Homework 1

Bayesian Machine Learning

1.

Your friend is on a gameshow and phones you for advice. She describes her situation as follows: There are three doors with a prize behind one of the doors and nothing behind the other two. She randomly picks one of the doors, but before opening it, the gameshow host opens one of the other two doors to show that it contains no prize. She wants to know whether she should stay with her original selection or switch doors. What is your suggestion? Calculate the relevant posterior probabilities to convince her that she should follow your advice.

Let W = i to be the event that the prize is behind door i, C = i to be the event that my friend chose door i, S = i the event that we are shown door i. Assume that we chose door 1,and then we are shown door 2 or 3. Then the probability, that we chose the winning door is:

$$P(W=1 \mid C=1, S=2) = \frac{P(S=2 \mid W=1, C=1) \cdot P(W=1, C=1)}{P(C=1, S=2)} =$$

$$= \frac{P(S=2 \mid W=1, C=1) \cdot P(W=1, C=1)}{P(S=2 \mid C=1, W=1) P(C=1, W=1) + P(S=2 \mid C=1, W=2) P(C=1, W=2) + P(S=2 \mid C=1, W=3) P(C=1, W=3)}$$

$$= \frac{1/2 \cdot 1/3}{1/2 \cdot 1/3 + 0 \cdot 1/3 + 1 \cdot 1/3} = \frac{1}{3}$$

That means if we stick with our initial choice we have a 1/3 chamce of winning, therefore we have to switch, and have a 2/3 chance of winning.

2.

Let $\pi = (\pi_1, \ldots, \pi_K)$, with $\pi_j \geq 0, \sum_j \pi_j = 1$. Let $X_i \sim \text{Multinomial}(\pi)$, i.i.d. for $i = 1, \ldots, N$. Find a conjugate prior for π and calculate its posterior distribution and identify it by name. What is the most obvious feature about the parameters of this posterior distribution?

The posterior distribution is:

$$p(\pi \mid X) = \frac{p(X \mid \pi)\, p(\pi)}{\int p(X \mid \pi)\, p(\pi)\, d\pi}$$

X_i is i.i.d multinomial, therefore:

$$p(X \mid \pi) = \prod_{i=1}^{N} (x_i \mid \pi) = \prod_{i=1}^{k} \pi_i^{c_i} \qquad c_k = \sum_{i=1}^{N} \mathbb{1}(X_i = k)$$

A conjugate prior for the multinomial distribution is the Dirichlet distribution, which is the multivariate generalization of the Beta distribution.

$$\pi \sim \mathrm{Dir}(\alpha)$$

$$p(\pi \mid \alpha) = \frac{\Gamma\left(\sum_{i=1}^{k} \alpha_i\right)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \prod_{i=1}^{k} \pi_i^{\alpha_i - 1}$$

$$p(X \mid \pi) \cdot p(\pi) = \frac{\Gamma\left(\sum_{i=1}^{k} \alpha_i\right)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \cdot \prod_{i=1}^{k} \pi_i^{c_i + \alpha_i - 1}$$

$$P(\pi|X) = \frac{\frac{\Gamma\left(\sum_{i=1}^{k} \alpha_i\right)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \prod_{i=1}^{k} \pi_i^{c_i+\alpha_i-1}}{\int \frac{\Gamma\left(\sum_{i=1}^{k} \alpha_i\right)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \prod_{i=1}^{k} \pi_i^{c_i+\alpha_i-1} \, d\pi}$$

The posterior is a Derichlet distribution with updated parameters:

$$P(\pi|X) = \left(\prod_{i=1}^{k} \pi^{c_i+\alpha_i-1}\right) \cdot \frac{\Gamma\left(\sum_{i=1}^{k} \alpha_i + c_i\right)}{\prod_{i=1}^{k} \Gamma(\alpha_i + c_i)} \sim \mathrm{Dir}(\alpha + c)$$

3.

You are given a dataset $\{x_1, \ldots, x_N\}$, where each $x \in \mathbb{R}$. You model it as i.i.d. Normal(μ, λ^{-1}). Since you don't know the mean μ or precision λ, you model them as $\mu|\lambda \sim \text{Normal}(0, a\lambda^{-1})$ and $\lambda \sim \text{Gamma}(b, c)$. Note that the priors are not independent.

a) Using Bayes rule, calculate the posterior of μ and λ and identify the distributions.

b) Using the posterior, calculate the predictive distribution on a new observation,

$$p(x^*|x_1, \ldots, x_n) = \int_0^\infty \int_{-\infty}^\infty p(x^*|\mu, \lambda) p(\mu, \lambda|x_1, \ldots, x_N) d\mu d\lambda$$

a, The posterior distribution is:

$$P(\mu, \lambda|X) = \frac{P(X|\mu, \lambda) \cdot P(\mu, \lambda)}{P(X)} = \frac{P(X|\mu, \lambda) P(\mu|\lambda) P(\lambda)}{P(X)}$$

Calculate the joint prior:

$$p(\mu|\lambda) = \frac{1}{\sqrt{2\pi a \lambda^{-1}}} e^{-\frac{\mu^2}{2a\lambda^{-1}}}$$

$$p(\lambda) = \frac{c^b}{\Gamma(b)} \lambda^{b-1} e^{-c\lambda}$$

$$p(\mu, \lambda) = \frac{c^b}{\sqrt{2\pi a}\ \Gamma(b)} \cdot \lambda^{b-1/2} \cdot \exp\left\{-\frac{\mu^2 \lambda}{2a} - c\lambda\right\}$$

The likelihood term:

$$p(X|\mu, \lambda) = \prod_{i=1}^{N} p(x_i|\mu, \lambda) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi \lambda^{-1}}} \cdot e^{\frac{-(x_i-\mu)^2}{2\lambda^{-1}}}$$

$$= \left(\sqrt{\frac{\lambda}{2\pi}}\right)^N \cdot e^{\sum_{i=1}^{N} -\frac{(x_i-\mu)^2 \lambda}{2}}$$

$$\propto \lambda^{N/2} \cdot \exp\left\{-\frac{\lambda}{2} \sum_{i=1}^{N} (x_i - \mu)^2\right\}$$

The posterior distribution is:

$$p(\mu,\lambda|X) = \frac{p(X|\mu,\lambda)\,p(\mu,\lambda)}{p(X)}$$

$$\propto p(X|\mu,\lambda)\,p(\mu,\lambda)$$

$$\propto \lambda^{N/2} \exp\left\{-\frac{\lambda}{2}\sum_{i=1}^{N}(x_i-\mu)^2\right\} \cdot \lambda^{b-1/2} \cdot \exp\left\{\frac{-\mu^2\lambda}{2a} - c\lambda\right\}$$

$$\propto \lambda^{N/2+b-1/2} \exp\left\{-\frac{\lambda}{2}\sum_{i=1}^{N}(x_i-\mu)^2 - \frac{\lambda}{2a}\mu^2 - c\lambda\right\}$$

We can rewrite the summation into the following form:

$$\sum_{i=1}^{N}(x_i-\mu)^2 = \sum_{i=1}^{N}(x_i-\bar{x}+\bar{x}-\mu)^2 = \sum_{i=1}^{N}\left[(x_i-\bar{x})^2+(\bar{x}-\mu)^2+2(x_i-\bar{x})(\bar{x}-\mu)\right] =$$

$$= \sum_{i=1}^{N}\left[(x_i-\bar{x})^2+(\bar{x}-\mu)^2\right] = \sum_{i=1}^{N}\left[(x_i-\bar{x})^2\right] + N(\bar{x}-\mu)^2$$

Substituting it back we have two exponential. In the next step we have to complete the square of the μ term

$$p(\mu,\lambda|X) \propto \lambda^{N/2+b-1/2} \exp\left\{-\lambda\left[c+\frac{1}{2}\sum_{i=1}^{N}(x_i-\bar{x})^2\right]\right\} \exp\left\{-\frac{\lambda}{2}\left(N(\bar{x}-\mu)^2+\frac{\mu^2}{a}\right)\right\}$$

Completer square:

$$N(\bar{x}-\mu)^2 + \frac{\mu^2}{a} = \frac{1}{a}\mu^2 + N(\bar{x}^2 + 2\bar{x}\mu + \mu^2) =$$

$$= \mu^2\left(\frac{1}{a}+N\right) - 2N\bar{x}\mu + N\bar{x}^2$$

$$= \left(\frac{1}{a}+N\right)\left(\mu^2 - \frac{2N\bar{x}}{\frac{1}{a}+N}\mu\right) + N\bar{x}^2$$

$$= \left(\frac{1}{a}+N\right)\left[\left(\mu - \frac{N\bar{x}}{\frac{1}{a}+N}\right)^2 - \frac{N\bar{x}}{\frac{1}{a}+N}\right] + N\bar{x}^2$$

$$= \left(\frac{1}{a}+N\right)\left(\mu - \frac{N\bar{x}}{\frac{1}{a}+N}\right)^2 + \frac{(N/a)\bar{x}^2}{(N+1/a)}$$

Substituting it back:

$$p(\mu,\lambda|X)$$

$$\propto \lambda^{N/2+b-1}\exp\left\{-\lambda\left[c+\frac{1}{2}\sum_{i=1}^{N}(x_i-\bar{x})^2\right]\right\}\exp\left\{-\frac{\lambda}{2}\left[\left(\frac{1}{a}+N\right)\left(\mu-\frac{N\bar{x}}{\frac{1}{a}+N}\right)^2+\frac{N/a\ \bar{x}^2}{\frac{1}{a}+N}\right]\right\}$$

$$\propto \lambda^{N/2+b-1}\exp\left\{-\lambda\left[c+\frac{1}{2}\sum_{i=1}^{N}(x_i-\bar{x})^2+\frac{\frac{N}{a}+\bar{x}^2}{2(\frac{1}{a}+N)}\right]\right\}\exp\left\{-\frac{\lambda}{2}\left(\frac{1}{a}+N\right)\left(\mu-\frac{N\bar{x}}{\frac{1}{a}+N}\right)^2\right\}$$

After some rearrangement we can notice the Normal-Gamma distribution with the parameters:

$$P(\mu, \lambda | X) \sim \text{Normal gamma}\left(\mu, \lambda \,\middle|\, \frac{N\bar{x}}{N + \frac{1}{a}}, N + \frac{1}{a}, b + \frac{N}{2}, c + \frac{1}{2}\left(\sum_{i=1}^{N}(x_i - \bar{x})^2 + \frac{\frac{N}{a}\bar{x}^2}{\frac{1}{a} + N}\right)\right)$$

b, Predictive distribution:

$$p(x^* | X) = \int_0^\infty \int_0^\infty P(x^* | \mu, \lambda) P(\mu, \lambda | X) \, d\mu \, d\lambda$$

$$\propto \int_0^\infty \int_{-\infty}^\infty \lambda^{1/2} \exp\left\{-\frac{1}{2}(x - \mu)^2 \lambda\right\} \lambda^{\alpha_0 - \frac{1}{2}} e^{-\beta_0 \lambda} \cdot \exp\left\{-\frac{\lambda_0 \lambda}{2}(\mu - \mu_0)^2\right\} d\mu \, d\lambda$$

$$\propto \int_0^\infty \int_{-\infty}^\infty \lambda^{\alpha_0} \exp\left\{\frac{-\lambda}{2}(x^* - \mu)^2 - \beta_0 \lambda - \frac{\lambda_0 \lambda}{2}(\mu - \mu_0)^2\right\} d\mu \, d\lambda$$

$$\propto \int_0^\infty \int_{-\infty}^\infty \lambda^{\alpha_0} \exp\left\{\frac{-\lambda}{2}\left(x^{*2} - 2x^*\mu + \mu^2\right) - \beta_0 \lambda - \frac{\lambda_0 \lambda}{2}\left(\mu^2 - 2\mu\mu_0 + \mu_0^2\right)\right\} d\mu \, d\lambda$$

$$\propto \int_0^\infty \int_{-\infty}^\infty \lambda^{\alpha_0} \exp\left\{-\frac{\lambda}{2}\left((1 + \lambda_0)\mu^2 - 2(x^* + \lambda_0 \mu_0)\mu + x^{*2} + \lambda_0 \mu_0^2 + 2\beta_0\right)\right\} d\mu \, d\lambda$$

Calculating the Gaussian integral with respect to μ

$$\propto \int_0^\infty \lambda^{\alpha_0-\frac{1}{2}} \exp\left\{-\frac{\lambda}{2}\left(x^{*2}+\lambda_0\mu_0^2+2\beta_0-\frac{(x^*+\lambda_0\mu_0)^2}{1+\lambda_0}\right)\right\} d\lambda$$

Calculating the Gaussian integral with respect to λ and rearranging:

$$\propto \left(x^{*2}+\lambda_0\mu_0^2+2\beta_0-\frac{(x^*+\lambda_0\mu_0)^2}{1+\lambda_0}\right)^{-\alpha_0-1/2}$$

$$\propto \left(\frac{\lambda_0(x^*-\mu_0)^2}{1+\lambda_0}+2\beta_0\right)^{-(\alpha_0+1/2)}$$

$$\propto \left(\frac{\lambda_0(x^*-\mu_0)^2}{2\beta_0(1+\lambda_0)}+1\right)^{-(\alpha_0+1/2)}$$

This is the form of a student t distribution:

$$P(x^*|X) \propto \left(1+\frac{1}{2\alpha_0}\frac{(x^*-\mu_0)^2}{\frac{\beta_0(1+\lambda_0)}{\alpha_0\lambda_0}}\right)^{-(\alpha_0+1/2)}$$

$$t_{2\alpha_0}\left(x^* \,\middle|\, \mu_0, \left(\frac{\beta_0(1+\lambda_0)}{\alpha_0 \lambda_0}\right)^{1/2}\right)$$

4. Bayesian calssifier

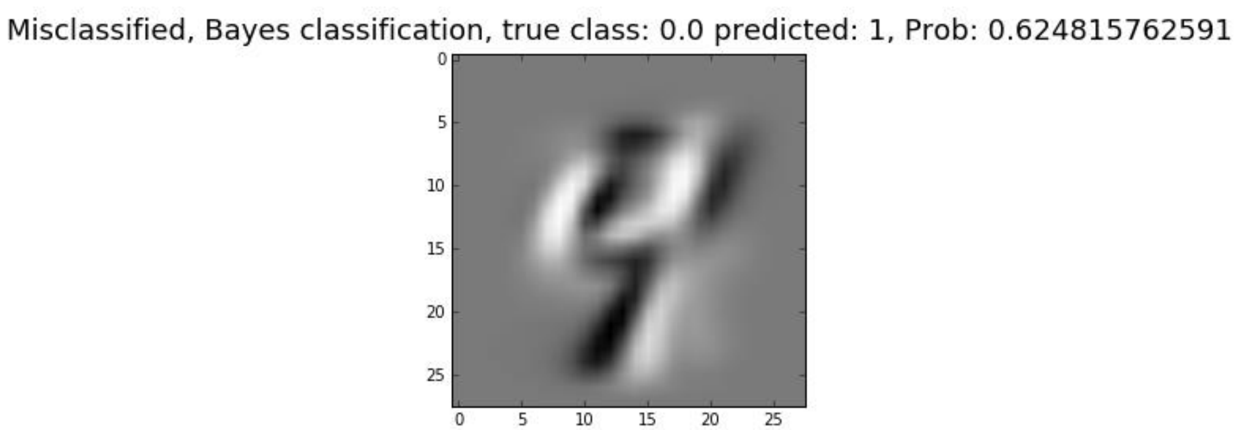
a, Code

b, Confusion matrix:

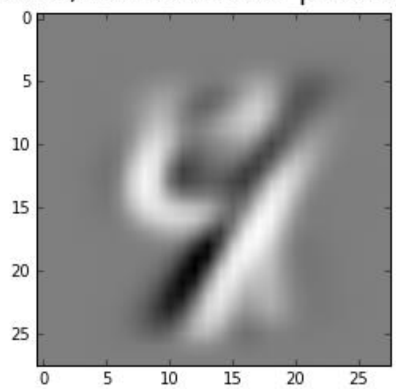
True Label	Predicted label 0	Predicted label 1
0	956	26
1	145	864

Accuracy: 0.914113510799

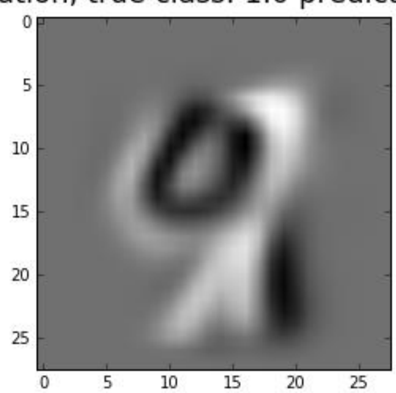
c, Misclassified images



Misclassified, Bayes classification, true class: 0.0 predicted: 1, Prob: 0.614349545987

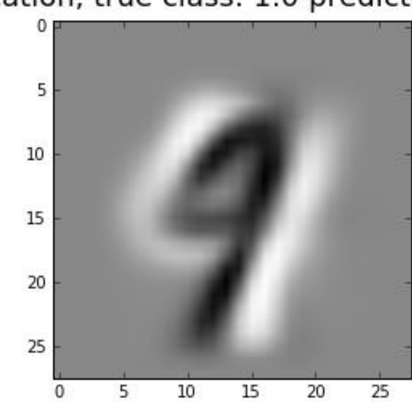


Misclassified, Bayes classification, true class: 1.0 predicted: 0, Prob: 0.636375309805

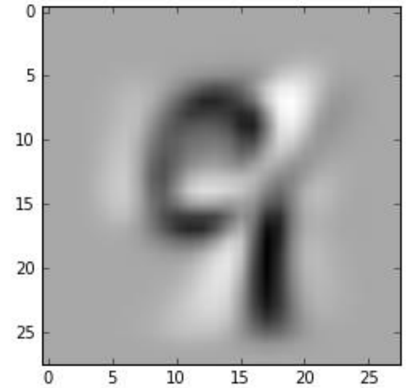


d, Ambiguous images:

Ambiguous, Bayes classification, true class: 1.0 predicted: 0, Prob: 0.500155088151



Ambiguous, Bayes classification, true class: 1.0 predicted: 1, Prob: 0.500825330038



Ambiguous, Bayes classification, true class: 0.0 predicted: 0, Prob: 0.501503575251