7.1

(1) $p(x|\omega_1,\omega_3,\omega_2,\omega_2) = 0.0076$

(2) $p(x|\omega_1,\omega_2,\omega_3,\omega_3) = 0.0125$

(3) $\omega_1,\omega_1,\omega_1,\omega_1$

7.2

13. If we choose the category ω_{max} that has the maximum posterior probability, our risk at a point $\mathbf{x}$ is:

$$\lambda_s \sum_{j \neq max} P(\omega_j|\mathbf{x}) = \lambda_s[1 - P(\omega_{max}|\mathbf{x})],$$

whereas if we reject, our risk is λ_r. If we choose a non-maximal category ω_k (where $k \neq max$), then our risk is

$$\lambda_s \sum_{j \neq k} P(\omega_j|\mathbf{x}) = \lambda_s[1 - P(\omega_k|\mathbf{x})] \geq \lambda_s[1 - P(\omega_{max}|\mathbf{x})].$$

This last inequality shows that we should never decide on a category other than the one that has the maximum posterior probability, as we know from our Bayes analysis. Consequently, we should either choose ω_{max} or we should reject, depending upon which is smaller: $\lambda_s[1 - P(\omega_{max}|\mathbf{x})]$ or λ_r. We reject if $\lambda_r \leq \lambda_s[1 - P(\omega_{max}|\mathbf{x})]$, that is, if $P(\omega_{max}|\mathbf{x}) \geq 1 - \lambda_r/\lambda_s$.

$\lambda_r = 0$时，一直拒绝，$\lambda_r > \lambda_s$，不会拒绝，退化为一般的最小风险决策。

7.3

(1)

$$P(x) = P(x = x_1, x_2, \cdots, x_{L-1})P(x_L|x_{L-1}) = P(x_2|x_1)P(x_3|x_2)\cdots P(x_L|x_{L-1}) = \Pi_1^{L-1} a_{x_{i+1}x_i}$$

(2)

$P(x|+) = \Pi_1^{L-1} a^+_{x_{i+1}x_i}, P(x|-) = \Pi_1^{L-1} a^-_{x_{i+1}x_i}$

所以:

$$S(x) = \log_2 \frac{P(x|+)}{P(x|-)} = \log_2 \frac{\Pi_i^{L-1} a^+_{x_{i+1}x_i}}{\Pi_i^{L-1} a^-_{x_{i+1}x_i}} = \sum_1^{L-1} (\log_2 a^+_{x_{i+1}x_i} - \log_2 a^-_{x_{i+1}x_i})$$

(3) 一共 27 条

8.1

$$\theta^* = \int_\theta \lambda(\hat{\theta}, \theta) p(\theta|X) d\theta = \int_\theta (\hat{\theta} - \theta)^2 p(\theta|X) d\theta$$

$$\frac{\partial \theta^*}{\partial \hat{\theta}} = \int_\theta 2(\hat{\theta} - \theta) p(\theta|X) d\theta = 0$$

$$\int_\theta \hat{\theta} p(\theta|X) d\theta = \int_\theta \theta p(\theta|X) d\theta = E(\theta|X)$$

$$\text{所以，当}\theta^*\text{最小时，}\theta^* = E(\theta|X) = \int_\theta \theta p(\theta|X) d\theta$$

8.2

(1)

由于$\max(L(\Theta)) \Leftrightarrow \max(\ln L(\Theta))$,所以考虑对$\ln L(\Theta)$最大化即可。

$$\ln L(\Theta) = \ln \prod_{i=1}^{n} \mathcal{N}(y^{(i)}|f_\Theta(x^{(i)});\sigma^2) = \sum_{i=1}^{n} \ln \mathcal{N}(y^{(i)}|f_\Theta(x^{(i)});\sigma^2) = \sum_{i=1}^{n} \ln(\frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{(f_\Theta(x^{(i)})-y^{(i)})^2}{2\sigma^2}))$$
$$= n\ln(\frac{1}{\sqrt{2\pi}\sigma}) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}((f_\Theta(x^{(i)})-y^{(i)})^2)$$

所以:

$$\max(L(\Theta)) \Leftrightarrow \max(\ln(L(\Theta))) \Leftrightarrow \max(n\ln(\frac{1}{\sqrt{2\pi}\sigma}) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}((f_\Theta(x^{(i)})-y^{(i)})^2)) \Leftrightarrow \min\sum_{i=1}^{n}((f_\Theta(x^{(i)})-y^{(i)})^2) \Leftrightarrow \min(E(\Theta))$$

(2)

$$\max p(\Theta|X,Y;\alpha^2,\sigma^2) \Leftrightarrow \max(p(Y|X;\Theta,\sigma^2)p(\Theta|\alpha^2)) \Leftrightarrow \max(\ln(p(Y|X;\Theta,\sigma^2)p(\Theta|\alpha^2)))$$
$$\Leftrightarrow \max\ln(\frac{1}{2\pi\sigma\alpha}\exp(-\frac{(f_\Theta(x^{(i)})-y^{(i)})^2}{2\sigma^2} - \frac{||\Theta||^2}{2\alpha^2})) \Leftrightarrow \min(-\frac{(f_\Theta(x^{(i)})-y^{(i)})^2}{2\sigma^2} - \frac{||\Theta||^2}{2\alpha^2})$$

当取$\lambda = \frac{1}{2\alpha^2}$时，即可知道两者等价。

添加L2正则相当于是给了一个先验概率，使得Θ的分布满足正态分布，从而对参数进行约束，因此相当于综合了参数的正态分布与ϵ的正态分布的综合影响，因此会降低过拟合

8.4

(1)

极大似然:

$$\hat{\mu} = \bar{x}, \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

贝叶斯、极大后验:

$$\hat{\mu} = \frac{n/\sigma^2}{n/\sigma^2 + 1/\sigma_0^2}\overline{x} + \frac{1/\sigma_0^2}{n/\sigma^2 + 1/\sigma_0^2}\mu_0$$