STA 414S/2104S: Some comments on HW 2

1. Solutions were generally fine, but some code is more elegant! In particular `R` lets you vectorize many calculations. Here is a nice way to discretize the wine data, due to Cody Severinksi:

```
wine$quality = as.factor(wine$quality)
levels(wine$quality) = c(rep("Bad", 2), "Acc", "Good", rep("Exc", 2))
```

   Most people found a classification error of around 40%, regardless of method, and often linear discriminant analysis was as good as anything.

   For the kernel discriminant analysis, it is not clear (to me) exactly where to incorporate the weights on each observation. But I think this is the most natural: define

$$\begin{aligned} \hat{\mu}_k^w &= \sum_{g_i=k} w_i^k x_i / N_k \\ \hat{\Sigma}^w &= \sum_{k=1}^{K} \sum_{g_i=k} w_i^k (x_i - \hat{\mu}_k^w)(x_i - \hat{\mu}_k^w)^T / (N-K) \end{aligned}$$

   where

$$w_i^k = \sum_{g_j=k} K_\lambda(x_i, x_j)$$

   possibly re-normalized to sum to 1. That is, the $\hat{\pi}$'s are left unchanged at N_k/N, and the observations in the kth group are weighted according to their distance from other observations in the kth group.

2. The solution is similar to that in HW 1, but you have to do a little more work yourself. This elegant solution is due to Viktoriya Krakovna. Writing

$$N(x)_{ij} = N_j(x_i), \quad \text{we have } y \sim \mathbb{N}(N(x)\theta, \sigma^2 I)$$

   making the usual assumption about ϵ. Write $\Sigma = \sigma^2\{N(x)^T N(x) + \lambda\Omega\}^{-1}$ and $\mu = \frac{1}{\sigma^2}\Sigma N^T(x) y$.

$$
\begin{aligned}
\pi(\theta \mid y) &= \frac{\pi(\theta) f(y \mid \theta)}{\int \pi(t) f(y \mid t) dt} \\
&= \frac{(2\pi)^{-N/2} |\frac{\sigma^2}{\lambda}\Omega^{-1}|^{-1/2} \exp(-\frac{1}{2}\theta^T \frac{\lambda}{\sigma^2}\Omega\theta)(2\pi)^{-N/2}\sigma^{-N} \exp(\frac{1}{2}(y - N\theta)^T \frac{1}{\sigma^2} I(y - N\theta))}{\int \{(2\pi)^{-N/2} |\frac{\sigma^2}{\lambda}\Omega^{-1}|^{-1/2} \exp(-\frac{1}{2}t^T \frac{\lambda}{\sigma^2}\Omega t)(2\pi)^{-N/2}\sigma^{-N} \exp(\frac{1}{2}(y - Nt)^T \frac{1}{\sigma^2} I(y - Nt))\} dt} \\
&= \frac{\exp[-\frac{1}{2}\{\frac{1}{\sigma^2} y^T y - \frac{2}{\sigma^2}\theta^T N^T(x) y + \frac{1}{\sigma^2}\theta^T (N(x)^T N(x) + \lambda\Omega)\theta\}]}{\int \exp[-\frac{1}{2}\{\frac{1}{\sigma^2} y^T y - \frac{2}{\sigma^2} t^T N^T(x) y + \frac{1}{\sigma^2} t^T (N(x)^T N(x) + \lambda\Omega) t\}] dt} \\
&= \frac{\exp\{-\frac{1}{2}(\frac{1}{\sigma^2} y^T y - 2\theta^T \Sigma^{-1} \frac{1}{\sigma^2}\Sigma N^T(x) y + \theta^T \Sigma^{-1}\theta)\}}{\int \exp\{-\frac{1}{2}(\frac{1}{\sigma^2} y^T y - 2\theta^T \Sigma^{-1} \frac{1}{\sigma^2}\Sigma N^T(x) y + t^T \Sigma^{-1} t)\} dt} \\
&= \frac{\exp[-\frac{1}{2}\{(\theta - \mu)^T \Sigma^{-1} (\theta - \mu) - \mu^T \Sigma^{-1}\mu + \frac{1}{\sigma^2} y^T y\}]}{\int \exp[-\frac{1}{2}\{(t - \mu)^T \Sigma^{-1} (t - \mu) - \mu^T \Sigma^{-1}\mu + \frac{1}{\sigma^2} y^T y\}] dt} \\
&= \frac{(2\pi)^{-N/2} |\Sigma|^{-1} \exp\{-\frac{1}{2}(\theta - \mu)^T \Sigma^{-1} (\theta - \mu)\} \exp\{-\frac{1}{2}(\mu^T \Sigma^{-1}\mu + \frac{1}{\sigma^2} y^T y)\}}{\int (2\pi)^{-N/2} |\Sigma|^{-1} \exp\{-\frac{1}{2}(t - \mu)^T \Sigma^{-1} (t - \mu)\} dt \exp\{-\frac{1}{2}(\mu^T \Sigma^{-1}\mu + \frac{1}{\sigma^2} y^T y)\}} \\
&= (2\pi)^{-N/2} |\Sigma|^{-1} \exp\{-\frac{1}{2}(\theta - \mu)^T \Sigma^{-1} (\theta - \mu)\}
\end{aligned}
$$

It follows directly that $E(N\theta \mid y) = N\mu = N(N^T N + \lambda\Omega)^{-1} N^T y$ and $\mathrm{cov}(N\theta \mid y) = \sigma^2 N(N^T N + \lambda\Omega)^{-1} N^T$.

Note that the prior needed to obtain this result is not very natural: $\theta \sim \mathbb{N}\{0, (\sigma^2/\lambda)\Omega^{-1}\}$, where the $N \times N$ matrix Ω is defined by

$$
\Omega_{ij} = \int N_i''(t) N_j''(t) dt.
$$