Statistical Learning: STA 7934 and CIS 6930
Exercises: Sections 4.1 - 4.3
Rebecca Steorts

1. $\sum_k \hat{Y}_k(\boldsymbol{x}) = 1$ for any $\boldsymbol{x}$ as long as there is an intercept in the model. We assume that $\boldsymbol{x}$ is a column vector.

   Proof:
$$\sum_k \hat{Y}_k(\boldsymbol{x}) = \mathbf{1}'_{\boldsymbol{K\times 1}}[(1, \boldsymbol{x})'\ \hat{B}]'.$$

$$= \mathbf{1}'_{\boldsymbol{K\times 1}}\hat{B}'(1, \boldsymbol{x}).$$
$$= [\hat{B}\ \mathbf{1}_{\boldsymbol{K\times 1}}]'(1, \boldsymbol{x}).$$

   Note that
$$\hat{B}\mathbf{1}_{\boldsymbol{K\times 1}} = (X'X)^{-1}X'Y\mathbf{1}_{\boldsymbol{K\times 1}}.$$

   Then
$$Y\mathbf{1}_{\boldsymbol{K\times 1}} = \mathbf{1}_{\boldsymbol{N\times 1}} \implies \hat{B}\mathbf{1}_{\boldsymbol{K\times 1}} = (X'X)^{-1}X\mathbf{1}_{\boldsymbol{N\times 1}}.$$

   Recall $(X'X)^{-1}(X'X) = I$. This implies that
$$(X'X)^{-1}X\mathbf{1}_{\boldsymbol{N\times 1}} = \boldsymbol{e_1}.$$

   Then
$$[\hat{B}\ \mathbf{1}_{\boldsymbol{K\times 1}}]'(1, \boldsymbol{x}) = \boldsymbol{e}'_{\mathbf{1}}(1, \boldsymbol{x}) = 1.$$

2. Fisher's problem amounts to maximizing
$$\max_{\boldsymbol{a}} \frac{\boldsymbol{a}'B\boldsymbol{a}}{\boldsymbol{a}'W\boldsymbol{a}}.$$

   Proof: Let $\boldsymbol{y} = W^{1/2}\boldsymbol{a}$. Then
$$\frac{\boldsymbol{a}'B\boldsymbol{a}}{\boldsymbol{a}'W\boldsymbol{a}} = \frac{\boldsymbol{y}'W^{-1/2}BW^{-1/2}\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{y}}.$$

   Consider
$$\frac{\boldsymbol{y}'W^{-1/2}BW^{-1/2}\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{y}} = \frac{\left(\frac{\boldsymbol{y}}{||\boldsymbol{y}||}\right)' W^{-1/2}BW^{-1/2}\left(\frac{\boldsymbol{y}}{||\boldsymbol{y}||}\right)}{\left(\frac{\boldsymbol{y}}{||\boldsymbol{y}||}\right)'\left(\frac{\boldsymbol{y}}{||\boldsymbol{y}||}\right)}.$$

   But $\left(\frac{\boldsymbol{y}}{||\boldsymbol{y}||}\right)'\left(\frac{\boldsymbol{y}}{||\boldsymbol{y}||}\right) = 1$ so without loss of generality, we can assume $\boldsymbol{y}'\boldsymbol{y} = 1$ and simply maximize

$$\boldsymbol{y}'W^{-1/2}BW^{-1/2}\boldsymbol{y}$$

subject to this constraint.

Let $A = W^{-1/2}BW^{-1/2}$ and note that it is symmetric. By the Spectral Decomposition Theorem, there exists an orthogonal matrix P and diagonal matrix D such that

$$\boldsymbol{y}'A\boldsymbol{y} = \boldsymbol{y}'PDP'\boldsymbol{y},$$

where $D = Diag(\lambda_1 \ldots \lambda_p)$ and P is composed of the eigenvectors of A.

Then let $x = P'y$ which implies that

$$\boldsymbol{y}'PDP'\boldsymbol{y} = \boldsymbol{x}'D\boldsymbol{x} = \sum_i \lambda_i x_i^2.$$

Thus,

$$\max_x \boldsymbol{y}'A\boldsymbol{y} = \max_{\boldsymbol{x}} \sum_i \lambda_i x_i^2.$$

Now consider

$$\sum_i \lambda_i x_i^2 \leq \sum_i \lambda_{max} x_i^2 = \lambda_{max}.$$

Note the last equality holds because $\boldsymbol{y}'\boldsymbol{y} = 1$ implies $\boldsymbol{x}'\boldsymbol{x} = 1$.

Suppose $\lambda_{max} = \lambda_k$ and take $y = e_k$. Then

$$\sum_i \lambda_{max} x_i^2 = \lambda_{max},$$

which implies given what we just showed that

$$\max_y \sum_i \lambda_i x_i^2 = \lambda_{max}.$$

By what we just showed above this implies that

$$\max_y \boldsymbol{y}'W^{-1/2}BW^{-1/2}\boldsymbol{y} = \max \text{eigenvalue}\{W^{-1/2}BW^{-1/2}\}$$

$$= \max \text{eigenvalue}\{W^{-1}B\}.$$