Ex 1

The "SVD" way is to note that a valid (thin) SVD of Q is

$$Q = \underset{U}{Q}\ \underset{\Sigma}{I_n}\ \underset{V^T}{I_n}$$

so that $Q^+ = I I Q^T = Q^T$. Then $\hat{x} = Q^T b$ as desired.
Alternatively, the normal equations give

$$Q^T Q \hat{x} = Q^T b$$

but $Q^T Q = I_n$, since Q has orthonormal columns, so this becomes

$$\hat{x} = Q^T b.$$

We haven't discussed computational complexity of LS problems, but it's lower when Q has orthonormal columns.

Ex 2

a) Start from the RHS:

$$(Ax)^T (A\hat{x}) = x^T A^T A \hat{x} \qquad \text{since } (Ax)^T = x^T A^T$$

$$= x^T A^T b \qquad \text{via the normal equations}$$

$$= (Ax)^T b \qquad \text{since } (Ax)^T = x^T A^T$$

b) Following the hint, we get

$$(A\hat{x})^T b = (A\hat{x})^T(A\hat{x}) = \|A\hat{x}\|^2$$

so that

$$\frac{(A\hat{x})^T b}{\|A\hat{x}\| \, \|b\|} = \frac{\|A\hat{x}\|^2}{\|A\hat{x}\| \, \|b\|} = \frac{\|A\hat{x}\|}{\|b\|}$$

c) Minimizing the angle between Ax and b is equivalent to maximizing the cosine of the angle, so instead we think about solving

$$\hat{x} = \underset{x}{\operatorname{argmax}} \frac{(Ax)^T b}{\|Ax\| \, \|b\|} = \underset{x}{\operatorname{argmax}} \frac{(Ax)^T (A\hat{x})}{\|Ax\| \, \|b\|} \quad \text{by part (a)}$$

Now note that by the Cauchy-Schwartz inequality,

$$(Ax)^T(A\hat{x}) = \langle Ax, A\hat{x} \rangle \leq \|Ax\| \, \|A\hat{x}\|$$

with equality when $Ax = \beta (A\hat{x})$ for some β, i.e., equality occurs when the angle between Ax and $A\hat{x}$ is zero. Therefore, setting $x = \beta \hat{x}$ maximizes $(Ax)^T(A\hat{x})$. By the remark, any value of β will minimize the angle, so we take $\beta = 1$.

Ex 3

Overdetermined means more equations/measurements than unknowns, which results in a unique LS solution. Underdetermined means A is wide and therefore has a nontrivial nullspace, so the solution is not unique. See pg. 4.24 for a summary.

Ex 4

In this case, we have

$$\Sigma = \begin{array}{rc} \begin{array}{r} r \\ \\ m-r \end{array} & \begin{bmatrix} \begin{matrix}\sigma_1 & & \\ & \ddots & \\ & & \sigma_r\end{matrix} & 0 \\ 0 & 0 \end{bmatrix} \\ & \begin{array}{cc} r & \qquad n-r \end{array} \end{array} \quad \text{and} \quad \Sigma^+ = \begin{array}{rc} \begin{array}{r} r \\ \\ n-r \end{array} & \begin{bmatrix} \begin{matrix}\frac{1}{\sigma_1} & & \\ & \ddots & \\ & & \frac{1}{\sigma_r}\end{matrix} & 0 \\ 0 & 0 \end{bmatrix} \\ & \begin{array}{cc} r & \qquad m-r \end{array} \end{array}$$

$$\Rightarrow \Sigma\Sigma^+ = \begin{array}{rc} \begin{array}{r} r \\ m-r \end{array} & \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \\ & \begin{array}{cc} r & m-r \end{array} \end{array}$$

Note the sizes!

Ex 5

For $K < \min(M,N)$, the rank of A is smaller/lower than $\min(M,N)$.

Ex 6

$$\left\| \begin{bmatrix} A \\ \sqrt{\beta} I \end{bmatrix} x - \begin{bmatrix} y \\ 0 \end{bmatrix} \right\|^2 = \left\| \begin{bmatrix} Ax - y \\ \sqrt{\beta} x - 0 \end{bmatrix} \right\|^2$$

$$= \begin{bmatrix} (Ax-y)^T & (\sqrt{\beta} x)^T \end{bmatrix} \begin{bmatrix} Ax - y \\ \sqrt{\beta} x \end{bmatrix}$$

$$= (Ax-y)^T (Ax-y) + \beta x^T x$$

$$= \|Ax-y\|^2 + \beta \|x\|^2$$

Ex 7

Let $A = V\Lambda V^{-1}$. Then

$$A^2 = V\Lambda V^{-1} V \Lambda V^{-1} = V \Lambda^2 V^{-1}$$

but if all eigenvalues are 0 or 1, then $\Lambda^2 = \Lambda$, so $A^2 = A$.

Ex 8

P is symmetric with eigenvalues all either 0 or 1, so we can write a thin eigenvalue decomposition as

$$P = V_r I_r V_r^T = Q Q^T$$

where $Q = V_r$ has r orthonormal columns.