1 Linear Regression, Projections and Pseudoinverses

(a) $|y-w|^2 = |y-P_X(y)+P_X(y)-w|^2 = |y-P_X(y)|^2 + |P_X(y)-w|^2 + 2(y-P_X(y))^\top(P_X(y)-w)$. Since $y-P_X(y)$ is orthogonal to any vector in $\mathrm{range}(X)$, we have $|y-w|^2 = |y-P_X(y)|^2 + |P_X(y)-w|^2$. It minimizes when $w = P_X(y)$. **QED.**

(b)

$\Leftarrow$ If $P = UU^\top$, then $P = P^T$ trivially holds. $P^2 = U(U^\top U)U^\top = UU^\top = P$. Since, $P = UU^\top$, we have $\mathrm{rank}(P) \le \mathrm{rank}(U) \le d$. $\mathrm{rank}(P) \ge \mathrm{rank}(U^\top PU) = \mathrm{rank}(I) = d$. Therefore, we have $\mathrm{rank}(P) = d$.

$\Rightarrow$ Since P is symmetric, we have $P = Q\Lambda Q^\top$, where $Q \in \mathbb{R}^{n\times n}$ is orthogonal and $\Lambda \in \mathbb{R}^{n\times n}$ is diagonal and real. Let λ be a eigenvalue of P and v be its corresponding eigenvector. Then we have $\lambda^2 v = P^2 v = Pv = \lambda v$. So the eigenvalues must be 0 or 1. Let U be the matrix of subset of columns of Q whose corresponding eigenvalues are 1. Since $\mathrm{rank}(P) = d$, there are d such columns. Since $P = \sum_{i=1}^n \lambda_i q_i q^\top$, we have $P = UU^\top$ with $U \in \mathbb{R}^{n\times d}$. Since U is orthogonal, we have $U^\top U = I$. **QED.**

(c) $\mathrm{tr}(P)$ is the sum of all the eigenvalues of P. we have proved in part (b) that all the eigenvalues are either 0 or 1, and since $\mathrm{rank}(P) = d$, we have $\mathrm{tr}(P) = d$.

(d) Let $X = U\Sigma V^\top$ be the SVD of X, where $U \in \mathbb{R}^{n\times d}, \Sigma \in \mathbb{R}^{d\times d}, V \in \mathbb{R}^{d\times d}$.

$$X(X^\top X)^{-1}X^\top = U\Sigma V^\top(V\Sigma U^\top U\Sigma V^\top)^{-1}V\Sigma U^\top = U\Sigma V^\top V(\Sigma^{-1})^2V^\top V\Sigma U^\top = UU^\top$$

This proves that $X(X^\top X)^{-1}X^\top$ is a rand-d orthogonal projection matrix. The corresponding matrix U is the matrix of the left singular vectors of X. **QED.**

(e)

Because the row space of X is the orthogonal complement of the null space of X. It is sufficient to show for any $v \in \mathbb{R}^n$, we have $Xv = 0 \Leftrightarrow \forall i, v_i^\top v = 0$.

$$Xv = \sum_{i:\sigma_i>0} \sigma_i u_i(v_i^\top v)$$

Since $\sigma_i u_i$'s are independent vectors, $Xv = 0 \Leftrightarrow \forall i, v_i^\top v = 0$. **QED.**

(f)

We can write X as the reduced form of the SVD: $X = U_r\Sigma_r V_r^T$. Then by definition of the Moore-Penrose pseudoinverse of X is $X^+ = V_r\Sigma_r^{-1}U_r^\top$. Then we have $X^+X = V_rV_r^T$, which is the orthogonal projection matrix onto the row space of X. If $\mathrm{rank}(X) = d$. Then $X^+ X = I$. If $\mathrm{rank}(X) = d$ and $n = d$, then $XX^+ = U_r U_r^\top = I$. So X^+ is the inverse of X, i.e. $X^+ = X^{-1}$.