Problem 1. Using Law of Total Probability:

$$\begin{aligned}
P(correct) &= P(correct|knew\ answer)P(knew\ answer) \\
&\quad + P(correct|eliminate\ choice)P(eliminate\ choice) \\
&\quad + P(correct|equally\ plausible)P(equally\ plausible) \\
&= 1 \cdot 0.5 + 0.25 \cdot 0.25 + 0.2 \cdot 0.25 \\
&= 0.6125
\end{aligned}$$

From Bayes' rule:

$$\begin{aligned}
P(knew\ answer|correct) &= \frac{P(correct|knew\ answer) \cdot P(knew\ answer)}{P(correct)} \\
&= \frac{1 \cdot 0.5}{0.6125} \\
&\approx \boxed{0.81632653061}
\end{aligned}$$

Problem 2. (a) Joint PMF:

$$\begin{aligned}
P(S=0, T=0) = P(X=0, Y=0) = 0.21 \\
P(S=1, T=1) = P(X=1, Y=0) = 0.09 \\
P(S=1, T=-1) = P(X=0, Y=1) = 0.49 \\
P(S=2, T=0) = P(X=1, Y=1) = 0.21
\end{aligned}$$

Marginal PMF of S: $P(S=0) = 0.21$, $P(S=1) = 0.58$, $P(S=2) = 0.21$.
Marginal PMF of T: $P(T=-1) = 0.49$, $P(T=0) = 0.42$, $P(T=1) = 0.09$.

(b) No, knowing the value of S gives us information about T.

(c)

$$E[S] = E[X+Y] = E[X] + E[Y] = 0.3 + 0.7 = 1$$

$$\text{Var}(2T) = 4\text{Var}(X-Y) = 4\text{Var}(X) - 4\text{Var}(Y) = 4(0.3)(0.7) + 4(0.7)(0.3) = 1.68$$

Problem 3. 1. $\text{rank}(X) = d$.

2. There is a unique solution for this system, since $\text{rank}(X) = d$ and there are d elements in w.

3. Since U and V are orthogonal:

$$\begin{aligned}
w &= X^\dagger y \\
&= X^T(XX^T)^{-1}y \\
&= V\Sigma^T U^T (U\Sigma V^T V \Sigma^T U^T)^{-1} y \\
&= V\Sigma^T U^T (U\Sigma\Sigma^T U^T)^{-1} y \\
&= V\Sigma^T U^T (U^T)^{-1} (\Sigma\Sigma^T)^{-1} U^{-1} y \\
&= V\Sigma^T (\Sigma\Sigma^T)^{-1} U^T y \\
&= V\Sigma^\dagger U^T y
\end{aligned}$$

4. The loss function is
$$\mathcal{L}(w) = \frac{1}{n}||Xw - y||^2$$
The gradient of the loss function is
$$\nabla_w \mathcal{L} = \frac{2}{n}\|Xw - y\|X$$
so the gradient descent iteration can be written as
$$w_t = w_{t-1} - \lambda\left(\frac{2}{n}\|Xw - y\|X\right)$$