Short Answer questions:

1. From Radon transform

$$\ln\left(\frac{I_0}{I}\right) = R_L\mu$$

The lefthand side can be obtained beforehand, the purpose of this transform is deriving μ.

Then using Fourier transform

$$F_\rho R_\mu(r,\varphi) = \int_{-\infty}^{+\infty} e^{-2\pi i r\rho} R_\mu(\rho,\varphi)d\rho = F\mu(\xi_1,\xi_2)$$

Then using inverse Fourier transformation, we can get $\mu(\xi_1,\xi_2)$.

2. The ramp filter does not permit low frequencies that cause blurring to appear in the image， but amplify the statistical noise present in the measured counts. **In order to reduce the amplification of high-frequencies,** the ramp filter is always combined with a low-pass filter, and in this case the low-pass filter is **Hamming window filter**

3.

	Advantage	Disadvantage
X-ray	Cheap, Low Radiation level	2D only
CT	3D appearance of human body. High contrast resolution and spatial resolution	High radiation level, Efficient
ultrasound	Cheap; fast and timely, no pain and danger, non-invasive. It has been popularized in clinical applications and it is an important part of medical imaging. Because the equipment is not as expensive as CT or MRI equipment, US diagnosis can obtain arbitrary cross-sectional images of organs, and can also observe the activity of moving organs.	the contrast resolution and spatial resolution of the image are not as high as that of CT or MRI.
MRI	Efficient; High resolution; Accurate	Expensive;

Calculation problems:

1.

a) The derivative of sigmoid function is:

$$\sigma'(x) = \frac{e^{-x}}{(1+e^{-x})^2}$$

Transform it into matrix form is:

$$\begin{bmatrix} \frac{e^{-x_1}}{(1+e^{-x_1})^2} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{e^{-x_k}}{(1+e^{-x_k})^2} \end{bmatrix}$$

b) The derivative of SoftMax function is

i. For $k = j$

$$\frac{\partial \mathrm{z_k}}{\partial x_j} = \frac{\exp(\mathrm{x}_k)}{\sum_{\mathrm{i=1}}^{\mathrm{K}} \exp(\mathrm{x_i})} (1 - \frac{\exp(\mathrm{x_k})}{\sum_{\mathrm{i=1}}^{\mathrm{K}} \exp(\mathrm{x_i})})$$

ii. For $k \neq j$

$$\frac{\partial \mathrm{z_k}}{\partial x_j} = -\frac{\exp(\mathrm{x_k})}{\sum_{\mathrm{i=1}}^{k} \exp(\mathrm{x_i})} \frac{\exp(\mathrm{x_j})}{\sum_{\mathrm{i=1}}^{j} \exp(\mathrm{x_i})})$$

Therefore, the matrix is:

$$\begin{bmatrix} \frac{\exp(\mathrm{x_1})}{\sum_{\mathrm{i=1}}^{1} \exp(\mathrm{x_i})} (1 - \frac{\exp(\mathrm{x_1})}{\sum_{\mathrm{i=1}}^{1} \exp(\mathrm{x_i})}) & \cdots & -\frac{\exp(\mathrm{x_k})}{\sum_{\mathrm{i=1}}^{k} \exp(\mathrm{x_i})} \frac{\exp(\mathrm{x_1})}{\sum_{\mathrm{i=1}}^{1} \exp(\mathrm{x_i})}) \\ 0 & \ddots & \vdots \\ 0 & 0 & \frac{\exp(\mathrm{x}_k)}{\sum_{\mathrm{i=1}}^{\mathrm{K}} \exp(\mathrm{x_i})} (1 - \frac{\exp(\mathrm{x_k})}{\sum_{\mathrm{i=1}}^{\mathrm{K}} \exp(\mathrm{x_i})}) \end{bmatrix}$$

2. The cross-entropy loss function is

$$L(\omega) = \frac{1}{N}\sum_i L_i = -\frac{1}{N}\sum_i \sum_{c=1}^{M} y_{ic}\log{(p_{ic})}$$

$$\frac{\partial L}{\partial \omega_n} = \frac{\partial L}{\partial p_j}\frac{\partial p_j}{\partial y_i}\frac{\partial y_i}{\partial \omega_n}$$

1) $\frac{\partial L}{\partial p_j} = -\sum_{j=1}^{K} \frac{y_i}{p_j}$

2) $\frac{\partial p_j}{\partial y_i} = \begin{cases} p_i\left(1 - p_i\right) & when\, j = i \\ -p_i p_j & when\, j \neq i \end{cases}$

3) $\frac{\partial y_i}{\partial \omega_n} = x_n$

Multiplex using 1), 2) and 3)

$$\frac{\partial L}{\partial \omega_n} = (-\sum_{i=1}^{K} y_i + p_i \sum_{i=1}^{K} y_i) x_n$$

In a classification case, the cross-entropy function represents the different between the true model and the current model. And the derivative of the cross-entropy function represents the updating rate of the process. The disadvantage of using naïve cross-entropy is when using naïve loss function (take MSE as an example), the learning rate will be very low at the beginning.