第一题

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

第二题

$\frac{df}{dx} = \frac{df}{dz} \cdot \frac{dz}{dy} \cdot \frac{dy}{dx}$

其中，$\frac{df}{dz} = -\frac{1}{2}e^{-\frac{1}{2}z}$, $\frac{dz}{dy} = y^T(S^{-1} + (S^{-1})^T)$, $\frac{dy}{dx} = I \in \mathbb{R}^{D\times D}$

$\therefore \frac{df}{dx} = -\frac{1}{2}e^{-\frac{1}{2}z}y^T(S^{-1} + (S^{-1})^T) = -\frac{1}{2}e^{-\frac{1}{2}((x-\mu)^T S^{-1}(x-\mu))}(x-\mu)^T(S^{-1} + (S^{-1})^T)$

偏导数维度：

$\frac{df}{dz} \in \mathbb{R}^{1\times 1}$, $\frac{dz}{dy} \in \mathbb{R}^{1\times D}$, $\frac{dy}{dx} \in \mathbb{R}^{D\times D}$, $\frac{df}{dx} \in \mathbb{R}^{1\times D}$

第三题

令 $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_D \end{pmatrix}$，则 $xx^T = \begin{pmatrix} x_1^2 & x_1x_2 & \cdots & x_1x_D \\ x_2x_1 & x_2^2 & \cdots & x_2x_D \\ \vdots & \vdots & \ddots & \vdots \\ x_Dx_1 & x_Dx_2 & \cdots & x_D^2 \end{pmatrix}$

因此，$f(x) = tr(xx^T + \sigma^2 I) = \sum_{i=1}^{D}(x_i^2 + \sigma^2)$

$\therefore \frac{df}{dx} = \begin{pmatrix} 2x_1 & 2x_2 & \cdots & 2x_D \end{pmatrix}$

第四题

设 $f : \mathbb{R}^n \to \mathbb{R}$ 在点 $\mathbf{x}$ 可微

取任意单位向量 $\mathbf{u}$，沿 $\mathbf{u}$ 方向的方向导数 $D_{\mathbf{u}}f(\mathbf{x}) = \nabla f(\mathbf{x}) \cdot \mathbf{u}$

$\because \nabla f(\mathbf{x}) \cdot \mathbf{u} \le \|\nabla f(\mathbf{x})\| \cdot \|\mathbf{u}\| = \|\nabla f(\mathbf{x})\|$ (当且仅当 $\mathbf{u}$ 与 $\nabla f(\mathbf{x})$ 共线且同向，即 $\mathbf{u} = \frac{\nabla f(\mathbf{x})}{\|\nabla f(\mathbf{x})\|}$, 取等号)

$\therefore$ 在所有单位方向中，方向导数的最大值为 $\|\nabla f(\mathbf{x})\|_2$，达到该最大值的方向就是梯度方向，故函数上升最快的方向为 $\nabla f(\mathbf{x})$；同理，下降最快的方向为 $-\nabla f(\mathbf{x})$

$\therefore$ 函数变化最快的方向是梯度方向，证明完毕

补充：若 $\nabla f(\mathbf{x}) = \mathbf{0}$，则对任意 $\mathbf{u}$ 有 $D_{\mathbf{u}}f(\mathbf{x}) = 0$，此时一阶变化对方向不敏感

第五题

$\because f(x) = x^TAx + b^Tx + c$

$\therefore \nabla f = x^T(A + A^T) + b^T$

$\therefore Hessian = \nabla^2 f = A + A^T$

第六题

$$f(x) = \langle Ax, x\rangle = (Ax)^T x = x^T A^T x = 4x_1^2 + 11x_1x_2 + 7x_2^2$$

令 $x_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

$$\therefore f(x_0) = 4 \cdot (1)^2 + 11 \cdot 1 \cdot 1 + 7 \cdot (1)^2 = 22$$

$$\because \nabla f = \begin{pmatrix} 8x_1 + 11x_2 & 14x_2 + 11x_1 \end{pmatrix}$$

$$\therefore \nabla f(x_0) = \begin{pmatrix} 19 & 25 \end{pmatrix}$$

$$\because Hessian = \begin{pmatrix} 8 & 11 \\ 11 & 14 \end{pmatrix}$$

$$\therefore \nabla^2 f(x_0) = \begin{pmatrix} 8 & 11 \\ 11 & 14 \end{pmatrix}$$

$$\begin{aligned} \therefore f(x) &= f(x_0) + \nabla f(x_0)(x - x_0) + \tfrac{1}{2}(x - x_0)^T \nabla^2 f(x_0)(x - x_0) \\ &= 22 + 19(x_1 - 1) + 25(x_2 - 1) + 4(x_1 - 1)^2 + 11(x_1 - 1)(x_2 - 1) + 7(x_2 - 1)^2 \end{aligned}$$

第七题

$$f(x) = e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

$$\therefore f'(x) = e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \cdot \left(-\tfrac{1}{2\sigma^2}\right) \cdot 2(x - \mu) = -\tfrac{1}{\sigma^2}(x - \mu)e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$