Appendix - Gradient and Jacobian of multivariate functions

- We use thin/bold letters, $x, v/\mathbf{x}, \mathbf{v}$, for denoting scalar/vector inputs.
  - $\mathbf{x} \in \mathbb{R}^n \Rightarrow \mathbf{x} = (x_1, \ldots, x_n)^T$, where $x_i \in \mathbb{R}, i \leq n$
- We use lower/upper case letters, $f, g, h/F, G, H$, for denoting scalar/vector functions.
  - $F \in \mathbb{R}^m \Rightarrow F = (f_1, \ldots, f_m)^T$, where $f_i \in \mathbb{R}, i \leq m$

1 Definition of gradient and Jacobian

- Univariate functions
  - Scalar function
    - (Function) $f(x) : \mathbb{R} \to \mathbb{R}$
    - (Gradient) $\nabla f(x) = df(x)/dx = \lim_{h\to 0}\{f(x+h) - f(x)\}/h : \mathbb{R} \to \mathbb{R}$
    - (Example)

$$f(x) = ax \Rightarrow \nabla f(x) = a$$

  - (Column) Vector function
    - (Function) $F(x) = \big(f_1(x), \ldots, f_m(x)\big)^T : \mathbb{R} \to \mathbb{R}^m$ (a column vector)
    - (Jacobian) $\nabla F(x) = \big(df_1(x)/dx, \ldots, df_m(x)/dx\big) : \mathbb{R} \to \mathbb{R}^{1\times n}$ (a row vector)
    - (Example)

$$F(x) = (3x+1, 2x^2 - 1, x^3 - 2x)^T \Rightarrow \nabla F(x) = (3, 4x, 3x^2 - 2)$$

→ 변수는 x 1개,
함수가 여러개.

$$F(x) = x\mathbf{a} = (a_1 x, \ldots, a_m x)^T \Rightarrow \nabla F(x) = \mathbf{a}^T$$

$$F(x) = g(x)\mathbf{a} = (a_1 g(x), \ldots, a_m g(x))^T \Rightarrow \nabla F(x) = \nabla g(x)\mathbf{a}^T$$

- Multivariate functions
  - Scalar function
    - (Function) $f(\mathbf{x}) = f\big(x_1, \ldots, x_n\big) : \mathbb{R}^n \to \mathbb{R}$
    - (Gradient) $\nabla f(\mathbf{x}) = \big(\partial f(\mathbf{x})/\partial x_1, \ldots, \partial f(\mathbf{x})/\partial x_n\big)^T : \mathbb{R}^n \to \mathbb{R}^n$ (a column vector)
    - (Example)

$$f(\mathbf{x}) = 3x_1^2 + 4x_2^2 + 5x_1x_2 \Rightarrow \nabla f(\mathbf{x}) = (6x_1 + 5x_2, 8x_2 + 5x_1)^T$$

$$f(\mathbf{x}) = \mathbf{a}^T\mathbf{x} = \mathbf{x}^T\mathbf{a} = a_1x_1 + \ldots + a_nx_n = \sum_{i=1}^n a_ix_i \Rightarrow \nabla f(\mathbf{x}) = \mathbf{a}$$

$$f(\mathbf{x}) = \mathbf{a}^T G(\mathbf{x}) = G(\mathbf{x})^T\mathbf{a} = a_1g_1(\mathbf{x}) + \ldots + a_mg_m(\mathbf{x}) = \sum_{j=1}^m a_jg_j(\mathbf{x}) \Rightarrow \nabla f(\mathbf{x}) = ???$$

  - Vector function
    - (Function) $F(\mathbf{x}) = \big(f_1(\mathbf{x}), \ldots, f_m(\mathbf{x})\big)^T : \mathbb{R}^n \to \mathbb{R}^m$
    - (Jacobian) $\nabla F(\mathbf{x}) = \big(\nabla f_1(\mathbf{x}), \ldots, \nabla f_m(\mathbf{x})\big) : \mathbb{R}^n \to \mathbb{R}^{n\times m}$ (a matrix)
    - (Example)

$$F(\mathbf{x}) = (2x_1x_2 + x_2^2, x_1^2 + 2x_1x_2)^T \Rightarrow \nabla F(\mathbf{x}) = \begin{pmatrix} 2x_2 & 2x_1 + 2x_2 \\ 2x_1 + 2x_2 & 2x_1 \end{pmatrix}$$

→ 변수가
여러개

$$F(\mathbf{x}) = \mathbf{A}\mathbf{x} = \begin{pmatrix} \mathbf{a}_1^T \\ \cdots \\ \mathbf{a}_m^T \end{pmatrix} \mathbf{x} = \begin{pmatrix} \mathbf{a}_1^T\mathbf{x} \\ \cdots \\ \mathbf{a}_m^T\mathbf{x} \end{pmatrix} \Rightarrow \nabla F(\mathbf{x}) = (\mathbf{a}_1, \cdots, \mathbf{a}_m) = \mathbf{A}^T$$

2 Gradient of products and ratios

- Univariate functions
  - Product and ratio between two scalars - skipped
  - Product of a vector and a scalar = vector

    1
    - (Function) $F(x) = G(x)h(x) : \mathbb{R} \to \mathbb{R}^m$
    - (Jacobian) $\nabla F(x) = \nabla G(x)h(x) + \nabla h(x)G(x)^T : \mathbb{R} \to \mathbb{R}^{1\times m}$
    - (Example)

$$F(x) = (\mathbf{a}x)h(x) \ \Rightarrow\ \nabla F(x) = \mathbf{a}^T h(x) + \nabla h(x)(\mathbf{a}x)^T$$
$$F(x) = \mathbf{a}\{xh(x)\} \ \Rightarrow\ \nabla F(x) = \{h(x) + \nabla h(x)x\}\mathbf{a}^T$$

  - Ratio of a vector to a scalar = vector
    - (Function) $F(x) = G(x)/h(x) : \mathbb{R} \to \mathbb{R}^m$
    - (Jacobian) $\nabla F(x) = \{\nabla G(x)h(x) - \nabla h(x)G(x)^T\}/h(x)^2 : \mathbb{R} \to \mathbb{R}^{1\times m}$
    - (Example)

$$F(x) = (\mathbf{a}x)/h(x) \ \Rightarrow\ \nabla F(x) = \{\mathbf{a}^T h(x) - \nabla h(x)(\mathbf{a}x)^T\}/h(x)^2$$
$$F(x) = \mathbf{a}\{x/h(x)\} \ \Rightarrow\ \nabla F(x) = [\{h(x) - x\nabla h(x)\}/h(x)^2]\mathbf{a}^T$$

- Multivariate functions

  벡터를 앞에.

  - Product of two scalars = scalar
    - (Function) $f(\mathbf{x}) = g(\mathbf{x})h(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$
    - (Gradient) $\nabla f(\mathbf{x}) = \nabla g(\mathbf{x})h(\mathbf{x}) + \nabla h(\mathbf{x})g(\mathbf{x})^T = \nabla g(\mathbf{x})h(\mathbf{x}) + \nabla h(\mathbf{x})g(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}^n$
    - (Example)

$$f(\mathbf{x}) = (\mathbf{a}^T\mathbf{x})(\mathbf{c}^T\mathbf{x}) \ \Rightarrow\ \nabla f(\mathbf{x}) = \mathbf{a}(\mathbf{c}^T\mathbf{x}) + \mathbf{c}(\mathbf{a}^T\mathbf{x})$$

  - Ratio of a scalar to a scalar = scalar
    - (Function) $f(\mathbf{x}) = g(\mathbf{x})/h(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$
    - (Gradient) $\nabla f(\mathbf{x}) = \{\nabla g(\mathbf{x})h(\mathbf{x}) - \nabla h(\mathbf{x})g(\mathbf{x})^T\}/h(\mathbf{x})^2 = \{\nabla g(\mathbf{x})h(\mathbf{x}) - \nabla h(\mathbf{x})g(\mathbf{x})\}/h(\mathbf{x})^2 : \mathbb{R}^n \to \mathbb{R}^n$
    - (Example)

$$f(\mathbf{x}) = \mathbf{a}^T\mathbf{x}/\mathbf{c}^T\mathbf{x} \ \Rightarrow\ \nabla f(\mathbf{x}) = \{\mathbf{a}(\mathbf{c}^T\mathbf{x}) - \mathbf{c}(\mathbf{a}^T\mathbf{x})\}/(\mathbf{c}^T\mathbf{x})^2$$
$$f(\mathbf{x}) = 1/\mathbf{c}^T\mathbf{x} \ \Rightarrow\ \nabla f(\mathbf{x}) = \{\mathbf{0}(\mathbf{c}^T\mathbf{x}) - \mathbf{c}(1)\}/(\mathbf{c}^T\mathbf{x})^2 = -\mathbf{c}/(\mathbf{c}^T\mathbf{x})^2$$

  2
  - Product of a vector and a scalar = vector
    - (Function) $F(\mathbf{x}) = G(\mathbf{x})h(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}^m$
    - (Jacobian) $\nabla F(\mathbf{x}) = \nabla G(\mathbf{x})h(\mathbf{x}) + \nabla h(\mathbf{x})G(\mathbf{x})^T : \mathbb{R}^n \to \mathbb{R}^{n\times m}$
    - (Example)

$$F(\mathbf{x}) = \mathbf{A}\mathbf{x}(\mathbf{a}^T\mathbf{x}) \ \Rightarrow\ \nabla F(\mathbf{x}) = \mathbf{A}^T(\mathbf{a}^T\mathbf{x}) + \mathbf{a}(\mathbf{A}\mathbf{x})^T$$
$$F(\mathbf{x}) = \mathbf{c}(\mathbf{a}^T\mathbf{x}) \ \Rightarrow\ \nabla F(\mathbf{x}) = \mathbf{0}(\mathbf{a}^T\mathbf{x}) + \mathbf{a}\mathbf{c}^T$$
$$F(\mathbf{x}) = (\mathbf{c}\mathbf{a}^T)\mathbf{x} \ \Rightarrow\ \nabla F(\mathbf{x}) = (\mathbf{c}\mathbf{a}^T)^T$$

  - Ratio of a vector to a scalar = vector
    - (Function) $F(\mathbf{x}) = G(\mathbf{x})/h(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}^m$

* (Jacobian) $\mathbb{R}^n \to \mathbb{R}^{n\times m}$: $\nabla F(\mathbf{x}) = \{\nabla G(\mathbf{x})h(\mathbf{x}) - \nabla h(\mathbf{x})G(\mathbf{x})^T\}/h(\mathbf{x})^2$
* (Example)

$$F(\mathbf{x}) = \mathbf{Ax}/\mathbf{a}^T\mathbf{x} \;\Rightarrow\; \nabla F(\mathbf{x}) = \{\mathbf{A}^T(\mathbf{a}^T\mathbf{x}) - \mathbf{a}(\mathbf{Ax})^T\}/(\mathbf{a}^T\mathbf{x})^2$$
$$F(\mathbf{x}) = \mathbf{c}/\mathbf{a}^T\mathbf{x} \;\Rightarrow\; \nabla f(\mathbf{x}) = \{\mathbf{0}(\mathbf{a}^T\mathbf{x}) - \mathbf{ac}^T\}/(\mathbf{a}^T\mathbf{x})^2$$

- Inner product of two vectors = scalar $\to \begin{pmatrix} g_1 \\ \vdots \\ g_m \end{pmatrix}^T \cdot \begin{pmatrix} h_1 \\ \vdots \\ h_m \end{pmatrix}$
  * (Function) $f(\mathbf{x}) = G(\mathbf{x})^T H(\mathbf{x}) = \sum_{j=1}^m g_j(\mathbf{x})h_j(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ $\to \Sigma\, \nabla(g_j(x)\cdot h_j(x))$
  * (Gradient) $\nabla f(\mathbf{x}) = \nabla G(\mathbf{x})H(\mathbf{x}) + \nabla H(\mathbf{x})G(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$
  * (Example) $= \Sigma\{\nabla g_j(x)\, h_j(x) + \nabla h_j(x)\, g_j(x)^T\}$

$$f(\mathbf{x}) = (\mathbf{Ax})^T(\mathbf{Bx}) \;\Rightarrow\; \nabla f(\mathbf{x}) = \mathbf{A}^T(\mathbf{Bx}) + \mathbf{B}^T(\mathbf{Ax}) = (\mathbf{A}^T\mathbf{B} + \mathbf{B}^T\mathbf{A})\mathbf{x}$$
$$f(\mathbf{x}) = \mathbf{x}^T\mathbf{Cx} = (\mathbf{C}^T\mathbf{x})^T(\mathbf{Ix}) \;\Rightarrow\; \nabla f(\mathbf{x}) = \mathbf{C}(\mathbf{Ix}) + \mathbf{I}(\mathbf{C}^T\mathbf{x}) = (\mathbf{C} + \mathbf{C}^T)\mathbf{x}$$

$= \begin{pmatrix} \nabla g_1 \\ \vdots \\ \nabla g_m \end{pmatrix}^T \begin{pmatrix} h_1 \\ \vdots \\ h_m \end{pmatrix} +$

$= \nabla G \cdot H + \nabla H \cdot G$

$\begin{pmatrix} \nabla h_1 \\ \vdots \\ \nabla h_m \end{pmatrix}^T \begin{pmatrix} g_1 \\ \vdots \\ g_m \end{pmatrix}$

3 Chain rules

- Outer univariate scalar
  - Inner univariate scalar - skipped
  - Inner multivariate scalar
    * (Function) $f(\mathbf{x}) = g(h(\mathbf{x})) \Leftrightarrow f = g(t), t = h(\mathbf{x}) : \mathbb{R}^n(\to \mathbb{R}) \to \mathbb{R}$
    * (Gradient) $\nabla f(\mathbf{x}) = \nabla h(\mathbf{x})\nabla g(h(\mathbf{x})) : \mathbb{R}^n \to \mathbb{R}$
    * (Example)

$$f(\mathbf{x}) = (\mathbf{a}^T\mathbf{x})^2 \;\Rightarrow\; \nabla f(\mathbf{x}) = \mathbf{a}2(\mathbf{a}^T\mathbf{x})$$
$$f(\mathbf{x}) = (\mathbf{a}^T\mathbf{x})(\mathbf{a}^T\mathbf{x}) \;\Rightarrow\; \nabla f(\mathbf{x}) = \mathbf{a}(\mathbf{a}^T\mathbf{x}) + \mathbf{a}(\mathbf{a}^T\mathbf{x})$$
$$f(\mathbf{x}) = g(\mathbf{a}^T\mathbf{x}) \;\Rightarrow\; \nabla f(\mathbf{x}) = \mathbf{a}\nabla g(\mathbf{a}^T\mathbf{x})$$

- Outer univariate vector
  - Inner univariate scalar
    * (Function) $F(x) = G(h(x)) \Leftrightarrow F = G(v), v = h(x) : \mathbb{R} \to \mathbb{R} \to \mathbb{R}^m$
    * (Jacobian) $\nabla F(x) = \nabla h(x)\nabla G(h(x)) : \mathbb{R} \to \mathbb{R}^{1\times m}$
    * (Example)

$$F(x) = \mathbf{a}(x^2 + x) \;\Rightarrow\; \nabla F(x) = (2x+1)\mathbf{a}^T$$
$$F(x) = \mathbf{a}h(x) \;\Rightarrow\; \nabla F(\mathbf{x}) = \nabla h(x)\mathbf{a}^T$$

  - Inner multivariate scalar
    * (Function) $F(\mathbf{x}) = G(h(\mathbf{x})) \Leftrightarrow F = G(v), v = h(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R} \to \mathbb{R}^m$
    * (Gradient) $\nabla F(\mathbf{x}) = \nabla h(\mathbf{x})\nabla G(h(\mathbf{x})) : \mathbb{R}^n \to \mathbb{R}^{n\times m}$
    * (Example)

$$F(\mathbf{x}) = \mathbf{a}(\mathbf{c}^T\mathbf{x}) \;\Rightarrow\; \nabla F(\mathbf{x}) = \mathbf{ca}^T$$
$$F(\mathbf{x}) = (\mathbf{ac}^T)\mathbf{x} \;\Rightarrow\; \nabla F(\mathbf{x}) = \mathbf{ca}^T$$
$$F(\mathbf{x}) = \mathbf{a}(\mathbf{c}^T\mathbf{x}) \;\Rightarrow\; \nabla F(\mathbf{x}) = \mathbf{0}(\mathbf{c}^T\mathbf{x}) + \mathbf{ca}^T$$

- Outer multivariate scalar

– Inner univariate vector
  - (Function) $f(x) = g(H(x)) \;\Leftrightarrow\; f = g(\mathbf{v}), \mathbf{v} = H(x) : \mathbb{R} \to \mathbb{R}^m \to \mathbb{R}$
  - (Gradient) $\nabla f(x) = \nabla H(x) \nabla g(H(x)) : \mathbb{R} \to \mathbb{R}$
  - (Example)

$$f(x) = \mathbf{a}^T(\mathbf{c}x) \;\Rightarrow\; \nabla f(x) = \mathbf{c}^T\mathbf{a}$$
$$f(x) = (\mathbf{a}^T\mathbf{c})x \;\Rightarrow\; \nabla f(x) = (\mathbf{a}^T\mathbf{c})^T$$

– Inner multivariate vector
  - (Function) $f(\mathbf{x}) = g(H(\mathbf{x})) \;\Leftrightarrow\; f = g(\mathbf{v}), \mathbf{v} = H(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}^m \to \mathbb{R}$
  - (Gradient) $\nabla f(\mathbf{x}) = \nabla H(\mathbf{x}) \nabla g(H(\mathbf{x})) : \mathbb{R}^n \to \mathbb{R}^n$
  - (Example)

$$f(\mathbf{x}) = \mathbf{a}^T(\mathbf{C}\mathbf{x}) \;\Rightarrow\; \nabla f(\mathbf{x}) = \mathbf{C}^T\mathbf{a}$$
$$f(\mathbf{x}) = (\mathbf{a}^T\mathbf{C})\mathbf{x} \;\Rightarrow\; \nabla f(\mathbf{x}) = (\mathbf{a}^T\mathbf{C})^T$$

- Outer multivariate vector

– Inner univariate vector
  - (Function) $F(x) = G(H(x)) \;\Leftrightarrow\; F = G(\mathbf{v}), \mathbf{v} = H(x) : \mathbb{R} \to \mathbb{R}^m \to \mathbb{R}^\ell$
  - (Jacobian) $\nabla F(x) = \nabla H(x) \nabla G(H(x)) : \mathbb{R} \to \mathbb{R}^{1\times\ell}$
  - (Example)

$$F(x) = \mathbf{C}(\mathbf{a}x) \;\Rightarrow\; \nabla F(x) = \mathbf{a}^T\mathbf{C}^T$$
$$F(x) = (\mathbf{C}\mathbf{a})x \;\Rightarrow\; \nabla F(x) = (\mathbf{C}\mathbf{a})^T$$

– Inner multivariate vector
  - (Function) $F(\mathbf{x}) = G(H(\mathbf{x})) \;\Leftrightarrow\; F = G(\mathbf{v}), \mathbf{v} = H(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}^m \to \mathbb{R}^l$
  - (Jacobian) $\nabla F(\mathbf{x}) = \nabla H(\mathbf{x}) \nabla G(H(\mathbf{x})) : \mathbb{R}^n \to \mathbb{R}^l$
  - (Example)

$$F(\mathbf{x}) = \mathbf{A}(\mathbf{C}\mathbf{x}) \;\Rightarrow\; \nabla F(\mathbf{x}) = \mathbf{C}^T\mathbf{A}^T$$
$$F(\mathbf{x}) = (\mathbf{A}\mathbf{C})\mathbf{x} \;\Rightarrow\; \nabla F(\mathbf{x}) = (\mathbf{A}\mathbf{C})^T$$

4 Statistical Example: Multivariate analysis

- Linear regression
  – (Function) $L(\boldsymbol\beta) = \sum_{i=1}^n (y_i - \mathbf{x}_i^T\boldsymbol\beta)^2 \;\Leftrightarrow\; (\mathbf{y} - \mathbf{X}\boldsymbol\beta)^T(\mathbf{y} - \mathbf{X}\boldsymbol\beta) : \mathbb{R}^p \to \mathbb{R}$
  – (Gradient) $\nabla L(\boldsymbol\beta) = \sum_{i=1}^n \{-2\mathbf{x}_i(y_i - \mathbf{x}_i^T\boldsymbol\beta)\} \;\Leftrightarrow\; -2\mathbf{X}^T(\mathbf{y} - \mathbf{X}\boldsymbol\beta) : \mathbb{R}^p \to \mathbb{R}^p$
  – (Hessian: Jacobian of gradient): $\nabla^2 L(\boldsymbol\beta) = \sum_{i=1}^n 2\mathbf{x}_i\mathbf{x}_i^T \;\Leftrightarrow\; 2\mathbf{X}^T\mathbf{X} : \mathbb{R}^p \to \mathbb{R}^{p\times p}$

$x_i^T \quad (x_1 \cdots x_n)$

- Logistic regression
  – (Function) $L(\boldsymbol\beta) = \sum_{i=1}^n \{-y_i\mathbf{x}_i^T\boldsymbol\beta + \log(1 + \exp(\mathbf{x}_i^T\boldsymbol\beta))\} : \mathbb{R}^p \to \mathbb{R}$
  – (Gradient) $\mathbb{R}^p \to \mathbb{R}^p$

$$\begin{aligned}\nabla L(\boldsymbol\beta) &= \textstyle\sum_{i=1}^n \{-(y_i\mathbf{x}_i^T)^T + \mathbf{x}_i \exp(\mathbf{x}_i^T\boldsymbol\beta)/(1 + \exp(\mathbf{x}_i^T\boldsymbol\beta))\} \\ &= \textstyle\sum_{i=1}^n \big[-y_i + \exp(\mathbf{x}_i^T\boldsymbol\beta)/\{1 + \exp(\mathbf{x}_i^T\boldsymbol\beta)\}\big]\mathbf{x}_i\end{aligned}$$

$\dfrac{\exp(x_i^T\beta)}{\{1+\exp(x_i^T\beta)\}^2}\, x_i x_i^T$

$\left(\dfrac{x_i \exp(x_i^T\beta)\cdot(1+\cancel{\exp(x_i^T\beta)}) - \cancel{x_i(\exp(x_i^T\beta)}\exp(x_i^T\beta)}{\{1+\exp(x_i^T\beta)\}^2} \right| x_i$

– (Hessian) $\nabla^2 L(\boldsymbol{\beta}) = \sum_{i=1}^n \exp(\mathbf{x}_i^T\boldsymbol{\beta})/\{1+\exp(\mathbf{x}_i^T\boldsymbol{\beta})\}^2 \mathbf{x}_i\mathbf{x}_i^T : \mathbb{R}^p \to \mathbb{R}^{p\times p}$

- General scheme
  - (Function) $L(\boldsymbol{\beta}) = \sum_{i=1}^n \ell_i(\mathbf{x}_i^T\boldsymbol{\beta})$: $\mathbb{R}^p \to \mathbb{R}$
  - (Gradient) $\nabla L(\boldsymbol{\beta}) = \sum_{i=1}^n \mathbf{x}_i \nabla \ell_i(\mathbf{x}_i^T\boldsymbol{\beta})$: $\mathbb{R}^p \to \mathbb{R}^p$
  - (Hessian) $\nabla^2 L(\boldsymbol{\beta}) = \sum_{i=1}^n \mathbf{x}_i \nabla^2 \ell_i(\mathbf{x}_i^T\boldsymbol{\beta})\mathbf{x}_i^T$: $\mathbb{R}^p \to \mathbb{R}^{p\times p}$

?. ~~Gradient~~

$$\nabla^2 L(\beta) = \nabla \sum_{i=1}^{n} \left(\frac{e^{x_i\beta}}{1+e^{x_i\beta}}\right) x_i$$

→ x_i는 벡터

벡터는 어리면 [illegible]

$\nabla G \cdot H + \nabla H \cdot G^T$

$$= \nabla \sum_{i=1}^{n} x_i \left(\frac{e^{x_i\beta}}{1+e^{x_i\beta}}\right)$$

$$= \sum_{i=1}^{n} \nabla x_i \left(\frac{e^{x_i\beta}}{1+e^{x_i\beta}}\right) + \nabla\left(\frac{e^{x_i\beta}}{1+e^{x_i\beta}}\right) x_i^T$$

$$= \sum_{i=1}^{n} \frac{x_i \cdot \exp(x_i^T\beta)}{\{1+e^{x_i^T\beta}\}^2} \cdot x_i^T$$

$$\frac{e+1^{xb}}{1+e^{x\beta}} \cdot 2 \frac{1}{1+e^{x\beta}}$$