Deep Learning Quiz Solution

1. (a) T (b) T (c) T

2.

$\because$ A is real symmetric

$\therefore$ All eigenvalues $\lambda_i \in \mathbb{R}$, All eigenvectors $\{v_i\}$ can be chosen to be orthogonal to each other

$\because$ $\{v_i\}$ forms an orthogonal basis for $\mathbb{R}^d$

$\therefore \exists \{c_i\} \subseteq \mathbb{R} \text{ s.t. } x = \sum_{i=1}^{d} c_i\, v_i$

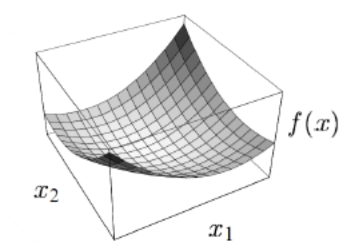
Thus, $x^{\mathrm{T}}\mathrm{A}x = x^{\mathrm{T}}\mathrm{A}(\sum_{i=1}^{d} c_i\, v_i) = x^{\mathrm{T}}(\sum_{i=1}^{d} c_i\, \mathrm{A}v_i) = x^{\mathrm{T}}(\sum_{i=1}^{d} c_i\, \lambda_i v_i)$

i. $x^{\mathrm{T}}\left(\sum_{i=1}^{d} c_i\, \lambda_i v_i\right) \le x^{\mathrm{T}}\lambda_{max} \sum_{i=1}^{d} c_i\, v_i = \lambda_{max} x^{\mathrm{T}}x \Rightarrow \frac{x^{\mathrm{T}}\mathrm{A}x}{x^{\mathrm{T}}x} \le \lambda_{max}, \forall x \neq 0$

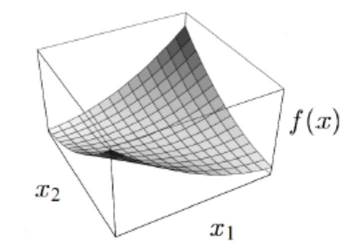
ii. $x^{\mathrm{T}}\left(\sum_{i=1}^{d} c_i\, \lambda_i v_i\right) \ge x^{\mathrm{T}}\lambda_{min} \sum_{i=1}^{d} c_i\, v_i = \lambda_{min} x^{\mathrm{T}}x \Rightarrow \frac{x^{\mathrm{T}}\mathrm{A}x}{x^{\mathrm{T}}x} \ge \lambda_{min}, \forall x \neq 0$

Hence, by i. and ii., $\lambda_{min} \le \frac{x^{\mathrm{T}}\mathrm{A}x}{x^{\mathrm{T}}x} \le \lambda_{max}$

3. (a)



(b)



4.

Solution 1:

$\mathrm{Var(y)} = \mathrm{Var}(\sum_i w_i x_i) = \mathrm{E}\left(\left(\sum_i \left(w_i x_i - \mathrm{E}(w_i x_i)\right)\right)^2\right) = \mathrm{E}\left(\sum_i \sum_j (w_i x_i - \mathrm{E}(w_i x_i)(w_j x_j - \mathrm{E}(w_j x_j))\right)$

$= \sum_i \sum_j w_i w_j Cov(x_i, x_j) = \sum_i \sum_j w_i w_j \Sigma_{ij} = \sum_i w_i (\sum_j w_j \Sigma_{ij}) = \sum_i w_i (\Sigma \quad w)_i = w^T \Sigma \quad w$

Solution 2:

$\mathrm{E(y)} = w^{\mathrm{T}}\mathrm{E}(x)$

$\mathrm{Var(y)} = \mathrm{E}\left((\mathrm{y} - \mathrm{E(y)})(\mathrm{y} - \mathrm{E(y)})^{\mathrm{T}}\right) = \mathrm{E}\left((w^{\mathrm{T}}x - w^{\mathrm{T}}\mathrm{E}(x))(w^{\mathrm{T}}x - w^{\mathrm{T}}\mathrm{E}(x))^{\mathrm{T}}\right)$

$= \mathrm{E}\left(w^{\mathrm{T}}\left(x - \mathrm{E}(x)\right)\left(x - \mathrm{E}(x)\right)^{\mathrm{T}} w\right) = w^{\mathrm{T}}\mathrm{E}\left(\left(x - \mathrm{E}(x)\right)\left(x - \mathrm{E}(x)\right)^{\mathrm{T}}\right) w$

$= w^{\mathrm{T}} \sum_x w$

5. Compared to the probability of disease given a symtom, P(y|x), it's easier to know the probability of a symptom given a disease, P(x|y). Accordingly, Bayes' rule simplify the procedure of evaluation.

6. 2

7. (a) Numeric issues may occur when $\mathrm{b} \gg \mathrm{a}$, because $\frac{a}{a+b}$ is near to 0 and $\log(0) \cong -\infty$.

(b) Convert $\log(\frac{a}{a+b})$ to $\log(\mathrm{a}) - \log(\mathrm{a} + \mathrm{b})$

8. (a) $\frac{\partial}{\partial \varepsilon} \mathrm{f}(\mathrm{a} + \varepsilon\vec{u}) = \nabla \mathrm{f}(\mathrm{a} + \varepsilon\vec{u})^{\mathrm{T}}\vec{u}$, when $\varepsilon = 0$, ans $= \nabla \mathrm{f}(\mathrm{a})^{\mathrm{T}}\vec{u}$

(b) $\mathrm{argmin}_{u,\|u\|=1} \nabla \mathrm{f}(\mathrm{a})^{\mathrm{T}}\vec{u} = \mathrm{argmin}_{u,\|u\|=1} \left\|\nabla \mathrm{f}(\mathrm{a})^{\mathrm{T}}\right\| \|\vec{u}\| cos\theta = \mathrm{argmin}_{u,\|u\|=1} cos\theta$

$\rightarrow \theta = 180^{\mathrm{o}}, u = -\nabla \mathrm{f}(\mathrm{a})^{\mathrm{T}}$