IA353 – Redes Neurais (1s2020)
Exercícios Conceituais 1 – EC 1
Gabarito das questões que envolvem manipulação algébrica

Questão 2)
Obtenção da solução ótima para o problema de minimização:

$$\|A\mathbf{x}-\mathbf{b}\|_P^2+\|\mathbf{x}-\mathbf{x}_0\|_Q^2$$

Resolução:

$$J(\mathbf{x})=\|A\mathbf{x}-\mathbf{b}\|_P^2+\|\mathbf{x}-\mathbf{x}_0\|_Q^2=(A\mathbf{x}-\mathbf{b})^T P(A\mathbf{x}-\mathbf{b})+(\mathbf{x}-\mathbf{x}_0)^T Q(\mathbf{x}-\mathbf{x}_0)$$

$$=\left(\mathbf{x}^T A^T-\mathbf{b}^T\right)(PA\mathbf{x}-P\mathbf{b})+\left(\mathbf{x}^T-\mathbf{x}_0^T\right)(Q\mathbf{x}-Q\mathbf{x}_0)$$

$$=\mathbf{x}^T A^T PA\mathbf{x}-\mathbf{x}^T A^T P\mathbf{b}-\mathbf{b}^T PA\mathbf{x}+\mathbf{b}^T P\mathbf{b}+\mathbf{x}^T Q\mathbf{x}-\mathbf{x}^T Q\mathbf{x}_0-\mathbf{x}_0^T Q\mathbf{x}+\mathbf{x}_0^T Q\mathbf{x}_0$$

$$\frac{\partial J(\mathbf{x})}{\partial \mathbf{x}}=2A^T PA\mathbf{x}-A^T P\mathbf{b}-A^T P\mathbf{b}+2Q\mathbf{x}-Q\mathbf{x}_0-Q\mathbf{x}_0=2A^T PA\mathbf{x}-2A^T P\mathbf{b}+2Q\mathbf{x}-2Q\mathbf{x}_0$$

Aplicando a condição necessária de otimalidade, resulta:

$$\left.\frac{\partial J(\mathbf{x})}{\partial \mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}^*}=0\Rightarrow A^T PA\mathbf{x}^*-A^T P\mathbf{b}+Q\mathbf{x}^*-Q\mathbf{x}_0=0$$

$$\left(A^T PA+Q\right)\mathbf{x}^*=A^T P\mathbf{b}+Q\mathbf{x}_0$$

As matrizes P e Q são sabidas serem simétricas, mas não são de nosso conhecimento. No entanto, a resposta desejada supõe a existência da inversa de $A^T PA+Q$, o que passaremos a considerar a partir daqui. Assim, multiplicando à esquerda por $\left(A^T PA+Q\right)^{-1}$ nos dois lados da equação acima, obtém-se:

$$\left(A^T PA+Q\right)^{-1}\left(A^T PA+Q\right)\mathbf{x}^*=\left(A^T PA+Q\right)^{-1}\left(A^T P\mathbf{b}+Q\mathbf{x}_0\right)$$

$$\mathbf{x}^*=\left(A^T PA+Q\right)^{-1}\left(A^T P\mathbf{b}+Q\mathbf{x}_0\right).$$

O coeficiente de regularização é fixo e igual a 1 porque a matriz Q já o inclui. Isso pode ser constatado de duas formas alternativas:

- Considere um coeficiente de regularização diferente de 1:

$$\|A\mathbf{x}-\mathbf{b}\|_P^2+\lambda\|\mathbf{x}-\mathbf{x}_0\|_Q^2$$

Essa nova formulação vai levar à solução:

$$\mathbf{x}^*=\left(A^T PA+\lambda Q\right)^{-1}\left(A^T P\mathbf{b}+\lambda Q\mathbf{x}_0\right)$$

Fazendo $Q'=\lambda Q$, retorna-se à formulação com coeficiente de regularização igual a 1:

$$\|A\mathbf{x}-\mathbf{b}\|_P^2+\|\mathbf{x}-\mathbf{x}_0\|_{Q'}^2$$

- Tomando $P = I$, $Q = \lambda I$ e $\mathbf{x}_0 = 0$, obtém-se a formulação original que estávamos acostumados a tratar, o que indica que o coeficiente de regularização de fato está embutido na matriz Q:

$$\|A\mathbf{x} - \mathbf{b}\|^2 + \lambda \|\mathbf{x}\|^2 .$$

Questão 3)

(a) Prove que matrizes $n \times n$ que podem ser decompostas na forma $M = N^T N$ com $N \in \Re^{m \times n}$, são matrizes simétricas e semi-definidas positivas.

Resolução:

$M^T = \left(N^T N\right)^T = N^T \left(N^T\right)^T = N^T N = M$. Logo, M é uma matriz simétrica.

Por definição, uma matriz M de dimensão $n \times n$ é semi-definida positiva se $\mathbf{x}^T M\mathbf{x} \geq 0, \forall \mathbf{x} \in \Re^n$. Usando a decomposição, resulta:

$$\mathbf{x}^T M\mathbf{x} = \mathbf{x}^T N^T N\mathbf{x} = (N\mathbf{x})^T (N\mathbf{x}) = \|N\mathbf{x}\|_2^2 .$$

Por definição, a norma euclidiana é tal que $\|\mathbf{z}\|_2 \geq 0, \forall \mathbf{z} \in \Re^m$, ou seja:

$$\mathbf{x}^T M\mathbf{x} = \|N\mathbf{x}\|_2^2 \geq 0, \forall \mathbf{x} \in \Re^n ,$$

demonstrando assim que M é semi-definida positiva.

(b) Supondo $m > n$ e tomando N de posto completo, prove que M tem que ser definida positiva.

Por definição, uma matriz M de dimensão $n \times n$ é definida positiva se $\mathbf{x}^T M\mathbf{x} > 0, \forall \mathbf{x} \in \Re_+^n$, onde $\Re_+^n$ representa o espaço $\Re^n$ sem o vetor nulo. Se a matriz N tem posto completo, então o seu espaço nulo tem dimensão zero, o que indica que somente o vetor nulo é solução de $N\mathbf{x} = 0$. Você pode recorrer também à definição de independência linear entre vetores e aplicar a $N\mathbf{x} = 0$, onde $N\mathbf{x}$ nada mais é que uma combinação linear das colunas de N. Logo:

$$\mathbf{x}^T M\mathbf{x} = \|N\mathbf{x}\|_2^2 > 0, \forall \mathbf{x} \in \Re_+^n,$$

demonstrando assim que M é definida positiva.

(c) Usando o conceito de matriz ortogonal, prove que N pode não ser única, para um mesmo M.

Resolução:

Por definição, uma matriz ortogonal Q é tal que $Q^T Q = I$, ou seja, $Q^T = Q^{-1}$. Sendo assim, tomando uma matriz Q ortogonal qualquer e de dimensão apropriada, sabe-se que $QN \neq N$. No entanto, se fizermos:

$\left(QN\right)^T\left(QN\right)=N^TQ^TQN=N^TIN=N^TN=M$. Logo, existe uma matriz diferente de N que produz M.

(d) Apresente uma proposta para obter N, dada uma matriz M definida positiva.

Resolução:

Sendo M uma matriz simétrica, é sempre possível obter n autovetores ortogonais para M, que irão compor as colunas de uma matriz T. Com essa matriz T, é possível chegar a uma matriz diagonal Λ que tenha os autovalores de M em sua diagonal, na forma:

$$M=T\Lambda T^T .$$

Como M é definida positiva, seus autovalores são todos positivos. Sendo assim, é possível obter uma matriz diagonal Λ_1 tal que os elementos da diagonal são a raiz quadrada dos autovalores de M. Sendo assim, obtém-se:

$$M=T\Lambda T^T=T\Lambda_1\Lambda_1T^T=T\Lambda_1^T\Lambda_1T^T=\left(\Lambda_1T^T\right)^T\left(\Lambda_1T^T\right).$$

Agora é só fazer $N=\Lambda_1T^T$, mostrando que é possível obter N a partir da decomposição espectral de M.

Uma alternativa para resolver esta questão seria recorrer à Fatoração de Cholesky (https://en.wikipedia.org/wiki/Cholesky_decomposition), mas aí teria que mostrar como fica a matriz N.

Questão 4)

Resolução:

Temos que lembrar que a linearização é sempre realizada em algum ponto específico, da mesma forma que o vetor gradiente (para o treinamento supervisionado de redes neurais) sempre é obtido em um ponto específico da superfície de erro. Com isso, a obtenção de $f_L(x,y)$, que é a versão linear de $f(x,y)=x\sqrt{y}$, requer expandir $f(x,y)$ em série de Taylor até primeira ordem, em torno de um determinado ponto (x_0,y_0). Procedendo dessa forma, no entorno de (x_0,y_0), resulta:

$$f(x,y)\cong f_L(x,y)=f(x_0,y_0)+\begin{bmatrix}\dfrac{\partial f(x,y)}{\partial x} & \dfrac{\partial f(x,y)}{\partial y}\end{bmatrix}_{(x,y)=(x_0,y_0)}\begin{bmatrix}x-x_0\\ y-y_0\end{bmatrix}$$

$$=x_0\sqrt{y_0}+\begin{bmatrix}\sqrt{y} & \dfrac{x}{2\sqrt{y}}\end{bmatrix}_{(x,y)=(x_0,y_0)}\begin{bmatrix}x-x_0\\ y-y_0\end{bmatrix}$$

$$=x_0\sqrt{y_0}+\sqrt{y_0}\left(x-x_0\right)+\frac{x_0}{2\sqrt{y_0}}\left(y-y_0\right)$$

$$=\sqrt{y_0}\,x+\frac{x_0}{2\sqrt{y_0}}\,y-\frac{x_0y_0}{2\sqrt{y_0}}$$

Logo:

$$f_L(x,y) = 2x + py - 8 = \sqrt{y_0}\,x + \frac{x_0}{2\sqrt{y_0}}\,y - \frac{x_0 y_0}{2\sqrt{y_0}},$$

o que conduz a:

$$\sqrt{y_0} = 2 \Rightarrow y_0 = 4$$

$$-\frac{x_0 y_0}{2\sqrt{y_0}} = -8 \Rightarrow x_0 = 8$$

$$\frac{x_0}{2\sqrt{y_0}} = p \Rightarrow p = 2$$

Cabe lembrar que é sempre possível conferir se o resultado está correto, ao comparar $f(x,y)$ com $f_L(x,y)$ no ponto (x_0, y_0).