Universidad Técnica Federico Santa María
Departamento de Informática

INF-393 Máquinas de Aprendizaje

PAUTA Quiz 2.

Se ruega al lector estudioso que comunique al profesor cualquier error que descubra en este documento.

1. *(25%) Verdadero o falso: "A un mayor valor de C, corresponde una mayor cantidad de vectores de soporte". Justifique.*

   Falso. En general, el número de support vectors aumenta al disminuir el valor de C. Esto ocurre porque con un menor valor de C, se permiten más errores de entrenamiento. De las condiciones de Karush-Kuhn-Tucker, observamos que estos datos necesariamente se convierten en support vectors, es decir, ejemplos $(x^{(\ell)}, y^{(\ell)})$ con multiplicadores $\alpha_{(\ell)} \neq 0$.

2. *(25%) ¿Cuál es el objetivo de la "poda" (pruning) en un árbol de clasificación?*

   Regularizar, es decir, evitar el sobre-ajuste. Adicionalmente se puede obtener un árbol que se puede evaluar de modo más eficiente computacionalmente.

3. *(25%) ¿Es cierto o es falso que una red neuronal de tipo feed-forward puede aproximar arbitrariamente bien cualquier función continua?¿Depende esto del número de capas? ¿Depende esto de la función de activación utilizada?*

   Es cierto si permitimos el uso de al menos una capa oculta y no limitamos el número de neuronas. Además, la función de activación debe ser de tipo squashing, aunque algunas versiones recientes del teorema de aproximación universal, permiten el uso de funciones no acotadas como rectificadoras lineales.

4. (50%) Si $K_1(x,z)$ y $K_2(x,z)$ son kernels admisibles

   (a) *(20%) Demuestre que $K(x,z) = K_1(x,z) + K_2(x,z)$ es un kernel admisible.*

   Si K_i es admisible $\exists \phi_i : K_i(x,z) = \langle \phi_i(x), \phi_i(z) \rangle$. Si consideramos $\phi(x) = (\phi_1(x), \phi_2(x))$, tenemos que

   $$\langle \phi(x), \phi(z) \rangle = \langle \phi_1(x), \phi_1(z) \rangle + \langle \phi_2(x), \phi_2(z) \rangle = K_1(x,z) + K_2(x,z) = K(x,z)\,, \tag{1}$$

   lo que implica que $\exists \phi : K(x,z) = \langle \phi(x), \phi(z) \rangle$, es decir, K es admisible.

   (b) *(30%) Demuestre que $K(x,z) = K_1(x,z)K_2(x,z)$ es un kernel admisible.*

Por el Teorema de Mercer, si K_i es admisible $\exists\{\psi_j^{(i)}\}_j, \exists\{\lambda_j^{(i)}\}_j : K_i(x,z) = \sum_j \lambda_j^{(i)} \psi_j^{(i)}(x)\psi_j^{(i)}(z)$. Por lo tanto

$$\begin{aligned} K(x,z) = K_1(x,z)K_2(x,z) &= \left(\sum_j \lambda_j^{(1)}\psi_j^{(1)}(x)\psi_j^{(1)}(z)\right)\left(\sum_k \lambda_k^{(2)}\psi_k^{(2)}(x)\psi_k^{(2)}(z)\right) \\ &= \sum\nolimits_j \sum\nolimits_k \lambda_j^{(1)}\lambda_k^{(2)}\psi_j^{(1)}(x)\psi_j^{(1)}(z)\psi_k^{(2)}(x)\psi_k^{(2)}(z) \\ &= \sum\nolimits_{i,k} \alpha_{jk}\psi_{jk}(x)\psi_{jk}(y)\,, \end{aligned} \tag{2}$$

lo que implica que $\exists\{\psi_i\}_i, \exists\{\alpha_i\}_i : K(x,z) = \sum_i \alpha_i\psi_i(x)\psi_i(z)$. Entonces, por el Teorema de Mercer, K es admisible.