Universidad Técnica Federico Santa María
Departamento de Informática

INF-393 Aprendizaje Automático

Cuestionario III 2018-II Campus San Joaquín

Los ejercicios marcados con (★) requieren un nivel mayor de profundización.

1. ¿En qué problemas consideraría el uso de redes convolucionales en vez de redes feed-forward clásicas?
2. Mencione 2 de las 3 ideas fundamentales que permiten reducir notablemente el número de parámetros entrenables de una capa convolucional con respecto a una capa densa tradicional.
3. A qué nos referimos con el término *mapa característico* (feature map) al describir la arquitectura de una red neuronal convolucional.
4. Explique conceptualmente la descomposición sesgo-varianza. Puede usar un ejemplo que ilustre la idea.
5. Verdadero o falso: *Bagging logra reducir el error de entrenamiento del learner base, exponencialmente rápido en el número de hipótesis ensambladas, independientemente de que éstas estén correlacionadas.* Justifique.
6. Explique la diferencia más importante entre Bagging y Random Forests (puede asumir que ambos usan árboles como learner base).
7. Verdadero o falso: *El muestreo de atributos que lleva a cabo Random Forests se realiza una sola vez, justo antes de entrenar al learner base.* Justifique.
8. ¿Qué son y cómo se construyen los *gráficos de importancia* con Random Forests? Podría usarse la misma técnica con Bagging o Adaboost.
9. Verdadero o falso: *Bagging intenta combinar hipótesis para reducir la varianza del learner base, tratando de mantener el sesgo invariable. Como consecuencia, es muy importante regularizar el entrenamiento de los predictores individuales.*. Justifique.
10. Verdadero o falso: *A diferencia de Bagging, Adaboost intenta combinar hipótesis para minimizar el sesgo del learner base, sin reducir activamente la varianza. Como consecuencia, es importante regularizar el entrenamiento de los predictores individuales.* Justifique.
11. ★ Considere el algoritmo Adaboost clásico (discreto) estudiado en clases. Sea $D_t(j)$ la distribución utilizada por el algoritmo para muestrear los ejemplos de entrenamiento al ajustar el t-ésimo modelo y ϵ_t el error obtenido, es decir $\epsilon_t = P_{D_t}(y^{(\ell)} \neq f_t(x^{(\ell)}))$. Demuestre que

    (a) Los pesos $\alpha_t = \ln(\epsilon_t/1-\epsilon_t)$ definidos por el algoritmo para construir la hipótesis ensamblada $F_t(x) = \text{sign}\left(\sum_j \alpha f_t(x)\right)$ son aquellos que minimizan el error de entrenamiento de $F_t(x)$,

    $$P_S(y^{(\ell)} \neq F_t(x^{(\ell)})) = \frac{1}{n}\sum_{\ell} I(y^{(\ell)} \neq F_t(x^{(\ell)}))\,. \tag{1}$$

    (b) Definiendo $\epsilon_t = 1/2 - \gamma_t$, error de la hipótesis ensamblada se puede acotar como

    $$P_S(y^{(\ell)} \neq F_t(x^{(\ell)})) \leq 2^t \prod_{j=1}^{t} \sqrt{1-4\gamma_t^2}\,. \tag{2}$$

12. Explique la diferencia más relevante entre Adaboost y Gradient Tree Boosting.

13. Verdadero o falso: *Para una función convexa* $g : \mathbb{R} \to \mathbb{R}$, $\mathbb{E}(g(X)) \geq g(\mathbb{E}(X))$.

14. ★ Demuestre que algoritmo EM es monótono, es decir que genera una secuencia de valores de los parámetros del modelo $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(t)}$, tal que $g(\theta^{(t)}) \leq g(\theta^{(t+1)})$, donde g es la función de verosimilitud correspondiente.

RNA+CVV LaTeX