Universidad Técnica Federico Santa María
Departamento de Informática

INF-578 Máquinas de Aprendizaje

Cuestionario Control 2 II 2017

1. Considere un punto bien clasificado, pero lejos de la frontera de decisión (margen funcional mayor que 1). ¿Por qué la frontera de decisión de la SVM lineal no se ve afectada por este punto, en cambio, la frontera de decisión de la regresión logística si se ve afectada?

2. Considere el problema primal de la C-SVM vista en clases. ¿En qué valor fijaría la constante C si el problema es linealmente separable?

3. En árboles de clasificación, ¿Cuándo se recomienda usar gain ratio sobre Information Gain?

4. En una C-SVM ¿Cuál es el rol del parámetro C?

5. Explique cuál es la estrategia implementada en árbol de clasificación para abordar problemas de clasificación que no son linealmente separables.

6. ¿En qué difieren los criterios de LDA y PCA para extraer características?

7. ([1] 4.2) Consideremos un problema de clasificación binaria con n datos $\mathbf{x}^{(\ell)} \in \mathbb{R}^d$, donde se tienen n_1 datos de la clase 1 y n_2 de la clase 2.

   (a) Muestre que la regla LDA clasifica la clase 2 si

   $$x^T\hat{\Sigma}^{-1}(\hat{\mu}_2 - \hat{\mu}_1) > \frac{1}{2}(\hat{\mu}_2 + \hat{\mu}_1)^T\hat{\Sigma}^{-1}(\hat{\mu}_2 - \hat{\mu}_1) - \log(n_2/n_1),$$

   y la clase 1 en caso contrario.

   (b) Suponga ahora que para cada dato $\mathbf{x}^{(\ell)}$ de clase 1 se define $y^{(\ell)} = -n/n_1$ y para cada dato de clase 2 se define $y^{(\ell)} = -n/n_2$ (es decir, las clases se codifican como $-n/n_1$ y $-n/n_2$ respectivamente). Considere luego la minimización de cuadrados:

   $$\sum_{\ell=1}^{n}(y^{(\ell)} - \beta_0 - \beta^T\mathbf{x}^{(\ell)})^2.$$

   Muestre que después de simplificaciones, la solución β satisface

   $$\left[(n-2)\hat{\Sigma} + n\hat{\Sigma}_B\right]\beta = n(\hat{\mu_2} + \hat{\mu_1}),$$

   donde $\hat{\Sigma}_B = \frac{n_1 n_2}{n^2}(\hat{\mu_2} + \hat{\mu_1})(\hat{\mu_2} + \hat{\mu_1})^T$

   (c) Continuando con el ejercicio anterior, muestre que $\hat{\Sigma}_B\beta$ está en la dirección $(\hat{\mu_2} - \hat{\mu_1})$, es decir

   $$\hat{\beta} \propto \hat{\Sigma}^{-1}(\hat{\mu_2} - \hat{\mu_1})$$

   Por lo tanto, el coeficiente de la regresión de de mínimos cuadrados es idéntico al coeficiente de LDA escalado por una constante.

8. ([1] 4.6) Supongamos que tenemos M puntos $\mathbf{x}_m \in \mathbb{R}^I$ con clases $\{-1, 1\}$. Pruebe que el algoritmo del perceptrón converge a un hiperplano separador en un número de pasos finito:

(a) Denote el hiperplano por $f(\mathbf{x}) = \beta^T x = 0$, (incluyendo el 1 en los vectores $\mathbf{x}$). Sea $z_m = \mathbf{x}_m/||\mathbf{x}_m||$. Muestre que separabilidad implica la existencia de un β_{sep} tal que $y_m \beta_{sep}^T z_m \geq 1, \forall m$.

(b) Dado un β^p en la iteración p, el perceptron identifica un punto z_m que está mal clasificado, y actualiza el vector $\beta^{p+1} = \beta^p + y_m z_m$. Muestre que $||\beta^{p+1} - \beta^{sep}||^2 \leq ||\beta^p - \beta^{sep}||^2 - 1$, por lo tanto el algoritmo converge al separar hiperplanos en no más que $||\beta^{inicial} - \beta^{sep}||^2$ pasos.

9. ([1] 4.7) Considere el criterio

$$D^*(\beta, \beta_0) = -\sum_{m=1}^{M} y_m(x_m^T \beta + \beta_0),$$

La suma de los márgenes funcionales sobre todo el training set. Considere maximizar D^* sujeto a $||\beta|| = 1$. Describa este criterio en palabras. ¿Este criterio resuelve el problema de encontrar el hiperplano separador óptimo?

References

[1] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., New York, NY, USA, 2009.

RNA+CVV LaTeX