Segundo examen parcial

Reconocimiento de Patrones (2024-2)

Julio Waissman Vilanova

Nombre: ____

1. Desarrolla un método de regularización para la regresión lineal basado en la varianza de los valores de los coeficientes w_j del modelo.

    - Escribe de manera clara cual sería la función a minimizar, y el pseudocódigo de entrenamiento utilizando un método simple de descenso de gradiente.
    - Explica qué ventajas o desventajas crees que presentaría este enfoque en comparación con la regularización l_2 (o regresión rígida) o l_1 (Lasso).

2. Desarrolla tu propio algoritmo de aprendizaje, utilizando Modelos Lineales Generalizados, pero asumiendo que los datos se presentan con una distribución de Poisson. En esta pregunta es más importante los procedimientos que los resultados, lo importante es ver que conocen y entienden las ideas generales para desarrollar Modelos Lineales Generalizados.

    a) Considera la distribución de Poisson, parametrizada como:

    $$\Pr(y; \lambda) = \frac{e^{-\lambda}\lambda^y}{y!}.$$

    Muestra que la distribución de Poisson pertenece a la familia exponencial, e indica claramente cuales son los valores de

    - η en función de λ,
    - $b(y)$,
    - $T(y)$,
    - $a(\eta)$,
    - λ en función de η,
    - $\hat{y} = E[y|x; \eta]$.

    b) Si asumimos un conjunto de aprendizaje $\{(x^{(1)}, y^{(1)}), \ldots, (x^{(M)}, y^{(M)})\}$, deriva el algoritmo de aprendizaje por descenso de gradiente, utilizando los logaritmos verosimilitudes lograrítmicas (loglikelyhood) de la misma forma que en clase las utilizamos para derivar el aprendizaje para la regresión lineal, logística y softmax. Escribe la formula final, y sobre todo todo el procedimiento que realizaste para llegar a ella.

3. Vuelve a realizar el mismo ejercicio, pero ahora asumiendo una distribución de Weibull, con $k > 0$ fijo. La distribución de Weibull tiene la siguiente forma:

$$\Pr(y; \lambda, k) = \begin{cases} \frac{k,\lambda}{(\frac{x,\lambda}{}}) ^{k-1} e^{-(x/\lambda)^k} & \text{si } x \geq 0, \\ 0 & \text{en otro caso.} \end{cases}$$

.

Recuerda que si seleccionas $k < 1$ la distribución es tipo Pareto, si $k = 1$ es exponencial y si $k = 2$ es una distribución de Rayleigh. Si asumes $1 < k < 2$ es una distribución de Weibull con forma de campana. Para cada caso el procedimiento será diferente, por lo que escoge uno solo, el que prefieras.

4. Propón y desarrolla una versión modificada de un clasificador SVM que permita tratar con datos altamente desbalanceados.

   En lugar de ajustar simplemente el parámetro de regularización C para cada clase, diseña una versión del SVM que modifique la función de margen de separación para dar un peso diferencial a los vectores de soporte de la clase minoritaria.

   Implementa este ajuste en el SVM y desarrolla el proceso para encontrar el problema de optiización dual. ¿En que es similar al visto en clase y en que difiere?

5. Desarrolla una variación del algoritmo de bosques aleatorios en la que cada árbol del bosque esté entrenado en un subconjunto aleatorio de las características (como en los bosques tradicionales), pero con una modificación adicional: en cada nodo de cada árbol, selecciona las características basándote en un criterio específico de importancia.

   Explica como sería en pseudocódigo, y como especificas el criterio de importancia. ¿Que diferencias tendría este algoritmo con el bosque aleatorio tradicional? ¿Qué ventajas o desventajas crees que tendría?

6. En un almacén, el equipo de logística está interesado en predecir la demanda diaria de ciertos productos. Sin embargo, la empresa incurre en diferentes costos dependiendo de si subestima o sobreestima la demanda.

   Si se subestima la demanda (es decir, se predice menos de lo que realmente se requiere), el costo es alto debido a la pérdida de ventas y la mala experiencia del cliente. Si se sobreestima la demanda (es decir, se predice más de lo necesario), el costo es menor, pero aún implica gastos adicionales en almacenamiento y manejo de inventarios.

   Para reflejar esta situación, se utiliza una función de pérdida asimétrica, definida de la siguiente forma:

$$loss(y, \hat{y}) = \begin{cases} \alpha|y - \hat{y}| & \text{si } \hat{y} < y, \\ \beta|y - \hat{y}| & \text{si } \hat{y} \geq y. \end{cases}$$

   donde y es la demanda real, $\hat{y}$ es la demanda predicha, y α y β son constantes que representan el costo de subestimación y sobreestimación, respectivamente.

   En este problema, tienes que aplicar el método de *gradient boosting* utilizando esta función de pérdida asimétrica.

   - Deriva el gradiente de la función de pérdida asimétrica con respecto a las predicciones $\hat{y}$ para *gradient boosting*. Nota que tendrás que considerar dos casos debido a la naturaleza asimétrica de la función de pérdida.
   - Describe cómo actualizarías las predicciones en cada iteración de *boosting* en función de este gradiente. Explica cómo este gradiente afecta la dirección de ajuste en cada paso.
   - ¿Qué se puede concluir sobre la utilidad de esta función de pérdida asimétrica en comparación con una función de pérdida simétrica como el error absoluto?