

Ödev #1

1 Ön bilgiler

Normal denklem için regularization kullanımı ve KDDCUP'99 saldırı tespit sistemleri için üretilen veri kümesinin lojistik regresyon modeli.

2 Sorular

1. $L2$ regularization için $(\lambda \|\mathbf{w}\|_2)$ analitik çözümün $\mathbf{w} = (X^T X + \lambda I)^{-1} X^T y$ olduğunu gösteriniz. (İpucu: kayıp fonksiyonuna ceza ($L2$) ekleyerek çözülebilir.) **Ek bilgiler:**

- $L2$ regularization için ifade : $\lambda \sum_{i=0}^n w_i^2 \Rightarrow \lambda \mathbf{w}^T \mathbf{w}$
- I ifadesi **Identity Matrix**.

$$I_n = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \in \mathbb{R}^{m \times m}$$

- $X^T X$ ifadesi $\mathbb{R}^{m \times m}$ 'dir ($X : \mathbb{R}^{m \times n}$ ve $X^T : \mathbb{R}^{n \times m} \Rightarrow X^T X \in \mathbb{R}^{m \times m}$). Bir matris ve skalar (λ) bir değer toplabilmesi için **identity matrix** ($I \in \mathbb{R}^{m \times m} \Rightarrow \lambda I$) ile çarpılması gerekir.

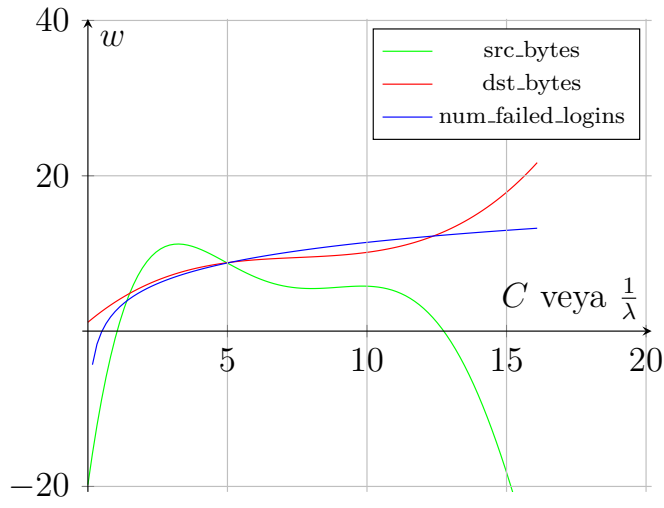
2. **Jupyter Notebook** üzerinde, KDDCUP99 verikümesi için 5 adet öznelik seçerek bunların λ değerlerini değiştirdiğinizde her bir öznelik çarpanının değişimini bir grafikte gösteriniz.

İpucu:

- `sklearn.linear_model.LogisticRegression` sınıfında yer alan **C** (Inverse of regularization strength) parametresi, regularization için kullanılan λ değerinin tersidir $C = \frac{1}{\lambda}$.
- Kodunuzu paralel hale getirerek daha hızlı çalışması amacıyla **n_jobs** parametresini çalışmasını istediğiniz CPU çekirdek sayısı olarak giriniz. Eğer bilgisayarınızda bulunan bütün çekirdeklerin kullanılması isterseniz **n_jobs = -1** olarak giriniz.

```
class sklearn.linear_model.LogisticRegression(penalty='l2', dual=False, tol=0.0001, C=1.0, fit_intercept=True, intercept_scaling=1, class_weight=None, random_state=None, solver='liblinear', max_iter=100, multi_class='ovr', verbose=0, warm_start=False, n_jobs=1)
```

Örnek Grafik



3 Notlandırma

1. sorunun çözümünü **tercihen pdf** (veya word) formatında, 2. sorunun çözümünü ise **Jupyter Notebook (*.ipynb)** dosyasını çalıştırarak çıktıları dosyada görünecek şekilde ozgur.catak@tubitak.gov.tr e-posta adresine gönderiniz.

Ödev son teslim tarihi 18.03.2018 23:59'dur. Bu tarihten sonra her gün için 25 puan kırılacaktır.