Hafta 14 - Mahremiyet Korumalı Makine Öğrenmesi

BGM 565 - Siber Güvenlik için Makine Öğrenme Yöntemleri

Bilgi Güvenliği Mühendisliği
Yüksek Lisans Programı

Dr. Ferhat Özgür Çatak
ozgur.catak@tubitak.gov.tr

İstanbul Şehir Üniversitesi
2018 - Bahar

İçindekiler

İçindekiler

Mahremiyet Nedir? I
Privacy

Mahremiyet

- Makine öğrenmesi, çeşitli kaynaklardan toplanan veri kümeleri kullanarak kullanıcı davranışları hakkında bilgi vermektedir.
- Bu yöntemler kullanılarak kişiler hakkında mahrem bilgilere ulaşmak mümkün olabilmektedir.
- Biyoenformatik, anayurt güvenliği, finans kayıtları

Mahremiyet Nedir? II
Privacy

Sweeney, 2002

- 2002 yılında Group Insurance Commission (GIC) tarafından yayınlanan **anonimleştirilmiş bir veri kümesi** ve seçmen kayıt listesi kullanılarak Massachusetts valisinin sağlık kayıtları ortaya çıkarıldı.

Voter registration list for Cambridge, Massachusetts

Name	Address	Date registered	Party affiliation	...	Zip	Birth date	Gender

Medical data in GIC

Ethnicity	Visit date	Diagnosis	Medication	...	Zip	Birth date	Gender

Mahremiyet Korumalı Makine Öğrenmesi I

Mahremiyet Korumalı Makine Öğrenmesi

- **Amaç**: **Veri kümesini** veya **modeli** yetkisiz kişilerin erişiminden korumak.
- **Yöntemler**
  - horizontal or vertical data distribution
  - data modification methods

	f_1	...	f_d
S_1			
...			
S_n			

Mahremiyet Korumalı Makine Öğrenmesi II

Partition

- Veri kümesinin yatay veya dikey olarak parçalanması
- **Yatay parçalama**: örneklerin parçalanması
- **Dikey parçalama**: Niteliklerin parçalanması

Mahremiyet Korumalı Makine Öğrenmesi III

Data Modification

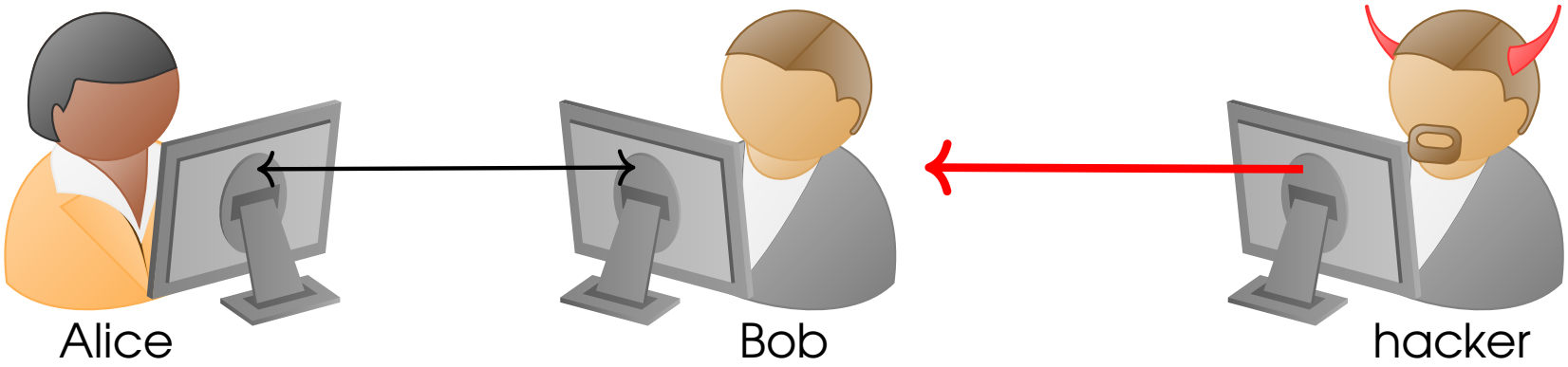
- **Perturbation**:
- **Blocking**: Bir niteliğin *NaN* ile değiştirilmesi
- **Birleştirme (Aggregation)**: Bir niteliği bir dağılım ile değiştirme
- **Swapping**: Örnekler arasında bilgilerin yerinin değiştirilmesi

Güvenli Çok Partili Hesaplama I
Secure Multi Party Computation

Multi-Parti Computation

- Genellikle kriptografik protokoller kullanılmaktadır.
- Diğer çalışmalar randomization/perturbation
- Dağıtık hesaplama (Distributed computation)
  - Veri kümesi farklı düğümler arasında dağıtılır.
  - Cooperative computations on private data

Güvenli Çok Partili Hesaplama II
Secure Multi Party Computation



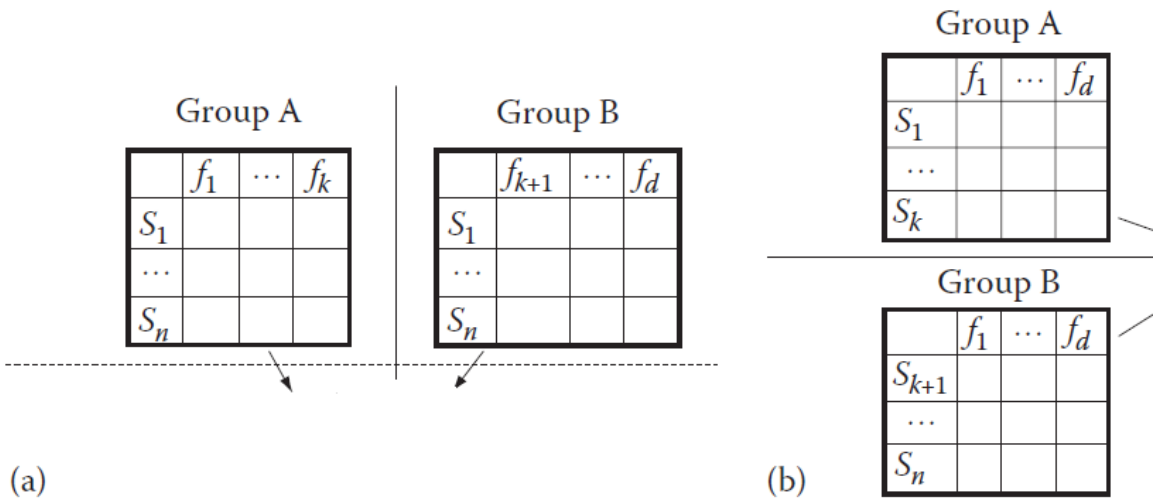
Örnek Senaryo

- **Alice**: Hacker profiles
- **Bob**: Sunucularına saldırı yapılmış olsun.
- **Bob** saldırı imzasını kullanarak **Alice**'in veritabanını kullanarak saldırganı tespit etmek istesin.
- Bunu yaparken **Alice**'den imza bilgisini gizlemek istiyor.
- Alice saldırganın mahrem bilgisini gizleyerek sadece ID bilgisini paylaşacak.

Cryptographic privacy preservation

MPC

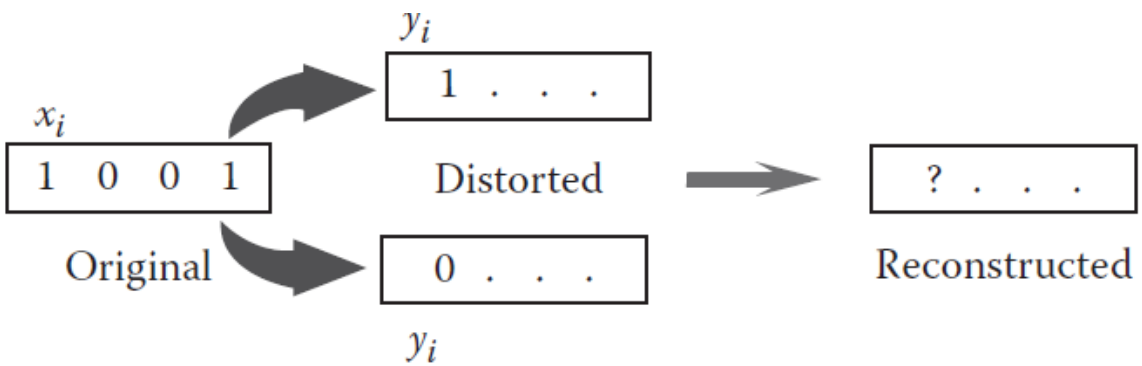
- Multiple participants: $P_1, \cdots, P_m$
- Compute a function: $f(DS_1, \cdots DS_m)$
- Her bir parti,P_i, DS verikümesinin sadece DS_i parçasına sahiptir.



Perturbation and Reconstruction

Perturbation

- Randomization kullanılarak hassas bilginin korunması
- Orjinal veri kümesi: $\mathcal{X} \in \mathbb{R}^{m \times n}$
- Perturbation matrix: $\mathcal{R} \in \mathbb{R}^{m \times n}$
- Distorted data set: $\mathcal{Y} = \mathcal{X} + \mathcal{R}$
- Reconstruction: $\mathcal{X} = \mathcal{Y} - \mathcal{R}$



Homomorphic Encryption I

Homomorphic Encryption

- Şifreli metinler üzerinde hesaplamaya izin veren bir şifreleme şeklidir.
- Sonuç şifrelenmeştir.
- Şifre çözüldüğünde, açık metin hesaplama ile aynı sonucu vermektedir.
- Homomorfik şifrelemenin amacı, şifrelenmiş veriler üzerinde hesaplamaya izin vermektir.
- Bulut bilişim için oldukça uygun bir hesaplama yöntemidir.

Homomorphic Encryption II

Partially homomorphic cryptosystems

- Kısmi Homomorfik Şifreleme (PHE) şemaları, belirli matematiksel işlemlerin şifreli veriler üzerinden hesaplanmasını sağlar.
- **Paillier cryptosystem** (additive homomorphic schemes):
  - $Enc(m_1) + Enc(m_2) = Enc(m_1 + m_2)$
  - $a \times m_1 = Dec(Enc(m_1)^a)$

Floating Point Numbers

- Paillier sadece tamsayılarla çalışabilmektedir.
- Veri kümesinden dönüşüm yapılmalıdır. $\mathbb{R}^{m \times n} \rightarrow \mathbb{Z}^{m \times n}$
- Gerçekleştirimlerde bir exponent ($K : 2^K$) çarpılarak rounding yapılır.
- Bilgi kaybı

Lab

Lab-1

İçindekiler

Privacy Preserving SVM I

SVM

- $$\min_{\alpha} \frac{1}{2}\alpha Q\alpha - \epsilon\alpha$$
  $$s.t.\ 0 \leq \alpha_i \leq v$$
  $$\sum_i d_i\alpha_i = 0 \quad i = 0, \cdots, m$$
- $X \in \mathbb{R}^{m \times n}$ $Q \in \mathbb{R}^{m \times m}$
- $Q_{ij} = K(\mathbf{x}_i, \mathbf{x}_j)$
  - **Linear**: $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \cdot \mathbf{x}_j$

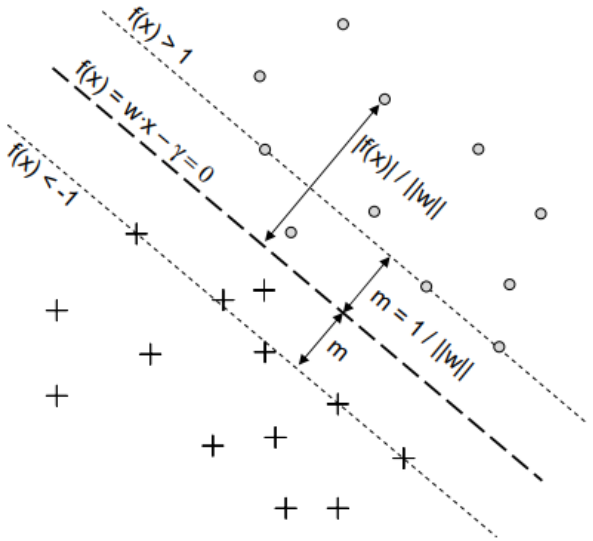


Figure 1: The separating hyperplane that maximizes the margin. ('o' is a positive data point, *i.e.* $f('o') > 0$, and '+' is a negative data point, *i.e.* $f('+') < 0$.)

Privacy Preserving SVM II

Kernel Matrix (Gram Matrix)

$\mathbf{K} = K(\mathbf{x}_i, \mathbf{x}_j), \ \textit{for } i, j = 1, \cdots, m$

$$K = \begin{bmatrix} K(\mathbf{x}_1, \mathbf{x}_1) & K(\mathbf{x}_1, \mathbf{x}_2) & K(\mathbf{x}_1, \mathbf{x}_3) & \dots & K(\mathbf{x}_1, \mathbf{x}_m) \\ K(\mathbf{x}_2, \mathbf{x}_1) & K(\mathbf{x}_2, \mathbf{x}_2) & K(\mathbf{x}_2, \mathbf{x}_3) & \dots & K(\mathbf{x}_2, \mathbf{x}_m) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ K(\mathbf{x}_m, \mathbf{x}_1) & K(\mathbf{x}_m, \mathbf{x}_2) & K(\mathbf{x}_m, \mathbf{x}_3) & \dots & K(\mathbf{x}_m, \mathbf{x}_m) \end{bmatrix}$$

Privacy Preserving SVM III

Yu, Hwanjo, Jaideep Vaidya, and Xiaoqian Jiang. "Privacy-preserving svm classification on vertically partitioned data." *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, Berlin, Heidelberg, 2006.

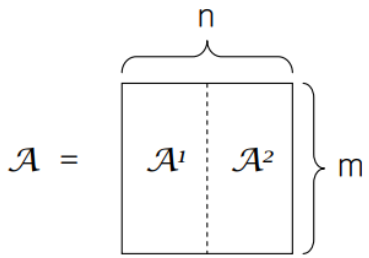


Fig. 3. Vertically partitioned matrix $\mathcal{A}$

Lemma 1. *Suppose the $m \times n$ data matrix $\mathcal{A}$ is vertically partitioned into $\mathcal{A}^1$ and $\mathcal{A}^2$ as Figure 3 illustrates. Let $\mathcal{K}^1$ and $\mathcal{K}^2$ be the $m \times m$ gram matrices of matrices $\mathcal{A}^1$ and $\mathcal{A}^2$ respectively. That is, $\mathcal{K}^1 = \mathcal{A}^1\mathcal{A}^{1'}$ and $\mathcal{K}^2 = \mathcal{A}^2\mathcal{A}^{2'}$. Then, $\mathcal{K}$, the gram matrix of $\mathcal{A}$, can be computed as follows:*

$$\mathcal{K} = \mathcal{K}^1 + \mathcal{K}^2 = \mathcal{A}^1\mathcal{A}^{1'} + \mathcal{A}^2\mathcal{A}^{2'} \tag{5}$$

Proof. An $(i,j)^{th}$ element of $\mathcal{K}$ is $x_i \cdot x_j$, where x_i and x_j are i^{th} and j^{th} data vectors in $\mathcal{A}$. Let x_i^1 and x_i^2 be vertically partitioned vectors of x_i, which are the parts from $\mathcal{A}^1$ and $\mathcal{A}^2$ respectively. Then,

$$x_i \cdot x_j = x_i^1 \cdot x_j^1 + x_i^2 \cdot x_j^2 \tag{6}$$

From Eq.(6), each element in $\mathcal{K}$ is equal to the sum of the elements in $\mathcal{K}^1$ and $\mathcal{K}^2$. Thus $\mathcal{K} = \mathcal{K}^1 + \mathcal{K}^2$.

Privacy Preserving SVM IV

Secure Merge of Local Models

- Oluşan *Gram matrislerin* ($Q \in \mathbb{R}^{m \times m}$) **güvenli** bir şekilde birleştirilmesi gerekmektedir.
- **Örnek protokol** (Sweeney and Shamos):
  - k adet parti ($P_0, P_1, \cdots, P_{k-1}$) ve bir parti ($P_i$) v_i şeklinde bir değer tutuyor olsun.
  - P_0, $\mathcal{F}$ uniform dağılımdan R elde eder. P_0 bunu lokal değere ekleyerek ($R + v_0 \ mod\ |\mathcal{F}|$) P_1'e gönderir.
  - Diğer partiler için şu şekilde devam eder.
    - P_i receives
    $$V = R + \sum_{j=0}^{i-1} v_j \ mod|\mathcal{F}|$$
    - P_i computes
    $$R + \sum_{j=0}^{i} v_j \ mod|\mathcal{F}| = (v_i + V) \ mod|\mathcal{F}|$$
  - P_0 kendine gelen değerden R değerini çıkardığı zaman toplam değere ulaşmış olur.

Lab

Lab-2

Homomorphic Encryption - Privacy Preserving SVM

Zhan, Justin, L. Chang, and Stan Matwin. "Privacy-preserving support vector machines learning." *Proceedings of the 2005 International Conference on Electronic Business (ICEB'05)*. 2005.

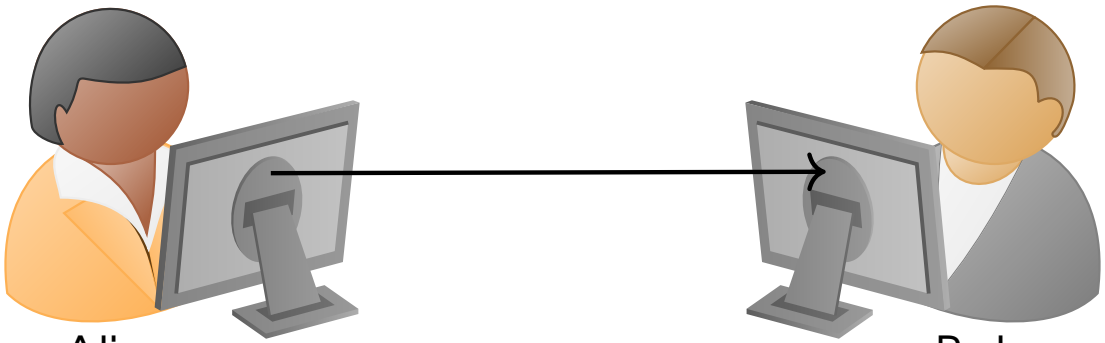
Protocol 1 *INPUT: P_1's input is a vector $\overrightarrow{x_1} = \{x_{11}, x_{12}, \cdots, x_{1m}\}$, and P_2's input is a vector $\overrightarrow{x_2} = \{x_{21}, x_{22}, \cdots, x_{2m}\}$. The elements in the input vectors are taken from the real number domain.*

1. P_1 performs the following operations:
   (a) She computes $e(x_{1i}+r_i)$s $(i \in [1,m])$ and sends them to P_2. r_i, known only by P_1, is a random number in real domain.
   (b) She computes $e(-r_i)$s $(i \in [1,m])$ and sends them to P_2.
2. P_2 performs the following operations:
   (a) He computes $t_1 = e(x_{11}+r_1)^{x_{21}} = e(x_{11} \cdot x_{21} + r_1 x_{21})$, $t_2 = e(x_{12}+r_2)^{x_{22}} = e(x_{12} \cdot x_{22} + r_2 x_{22})$, $\cdots$, $t_m = e(x_{1m})^{x_{2m}} = e(x_{1m} \cdot x_{2m} + r_m x_{2m})$.
   (b) He computes $t_1 \times t_2 \times \cdots \times t_m = e(x_{11} \cdot x_{21} + x_{12} \cdot x_{22} + \cdots + x_{1m} \cdot x_{2m} + r_1 x_{21} + r_2 x_{22} + \cdots + r_m x_{2m}) = e(\overrightarrow{x_1} \cdot \overrightarrow{x_2} + \sum_{i=1}^{m} r_i x_{2i})$.
   (c) He computes $e(-r_i)^{x_{2i}} = e(-r_i x_{2i})$ for $i \in [1,m]$.
   (d) He computes $e(\overrightarrow{x_1} \cdot \overrightarrow{x_2} + \sum_{i=1}^{m} r_i x_{2i}) \times e(-r_1 x_{21}) \times e(-r_2 x_{22}) \times \cdots \times e(-r_m x_{2m}) = e(\overrightarrow{x_1} \cdot \overrightarrow{x_2})$.

Lab

Lab-3

Logistic Regression Encrypted Model I



- Alice bir sınıflandırıcı eğitmektedir (spam classifier)
- Oluşan bu modeli Bob'un kişisel e-postalarına uygulamak istemektedir.
  - Alice eğitim kümesini ve modeli paylaşmamaktadır.
  - Bob e-postalarını paylaşmadan ve şifreli modeli kullanarak sınıf etiketlerini öğrenmektedir.

Logistic Regression Encrypted Model II

Adımlar

- Alice, spam e-posta veri kümesini kullanarak logistic regression ile bir model oluşturur.
- Alice, Paillier'i kullanarak public/private key pair oluşturur.
- Model public key ile şifrelenmiştir.
- Public key ve model Bob'a gönderilir.
- Bob, herbir e-posta için şifreli modeli kullanarak şifreli skorları elde eder.
- Bob şifreli skorları Alice'e gönderir.
- Alice şifreli skor değerleri açarak spam olup olmadığı bilgisini Bob'a geri gönderir.

Lab

Lab-4