Recitation

Lyle Ungar

Computer and information Science

Learning Objectives

PSD

Kernel and kernel matrix

Scale invariance

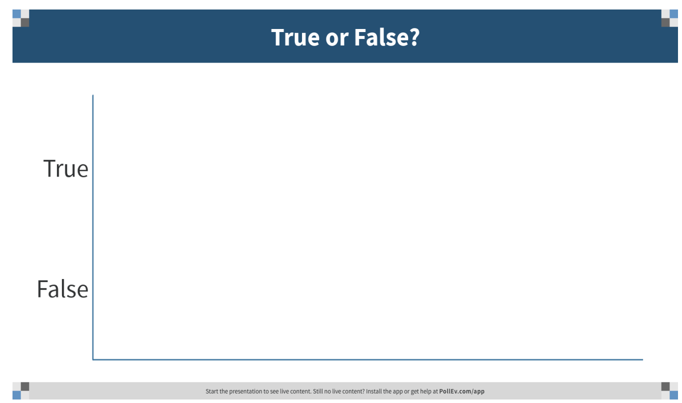
A kernel *k(x,y)*

- **Measures the *similarity* between a pair of points *x* and *y***
- **Symmetric and positive semi-definite**
- **Often tested using a *Kernel Matrix,***
  - a PSD matrix $\mathbf{K}$ with elements $K_{ij} = k(\mathbf{x}_i,\mathbf{x}_j)$ from all pairs of rows of a matrix X of predictors
  - A *PSD matrix* has only non-negative eigenvalues

Positive Semi-Definite (PSD)?

- $\begin{vmatrix} 1 & 2 \\ 2 & 1 \end{vmatrix}$ is positive semi-definite?
- A'A is guaranteed positive semi-definite?
- A positive semi-definite matrix can have negative entries in it?
- The covariance matrix is PSD?

True or False?



Example kernels

- **Linear kernel**
  - $k(x,y) = x^T y$
- **Gaussian kernel**
  - $k(x,y) = \exp(-||x - y||^2/\sigma^2)$
- **Quadratic kernel**
  - $k(x,y) = (x^T y)^2$ or $(x^T y + 1)^2$
- **Combinations and transformations of kernels**

Kernel matrix example

- Pick a matrix X

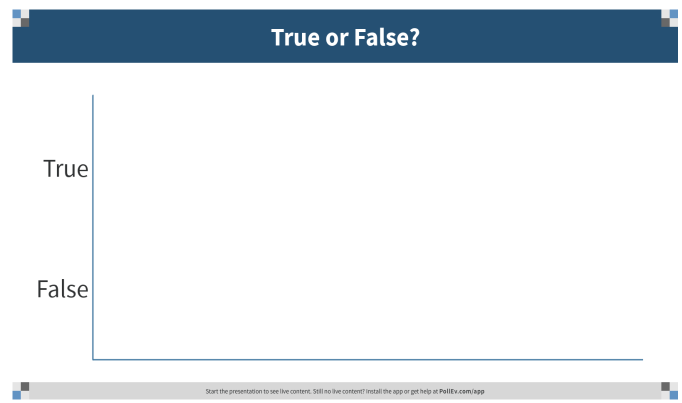
$$\begin{vmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{vmatrix}$$

- Compute $K_{ij} = k(\mathbf{x}_i,\mathbf{x}_j)$
- Test the eigenvalues

- What is K for X using the linear kernel?

True or false

- Kernels in effect transform observations **x** to a higher dimension space **ϕ(x)**
- Since kernels measure similarity,
  - k(**x**,**y**) < k(**x**,**x**) for **x** != **y**
- If there exists a pair of points **x** and **y** such that *k(**x**,**y**) < 0*, then *k()* is not a kernel



Kernels: True or false

- A quadratic kernel $(\mathbf{x}^T\mathbf{y})^2$, when used in linear regression, gives results very similar to including quadratic interaction terms in the regression
- Any distance metric $d(\mathbf{x},\mathbf{y})$ can be used to generate a kernel using $k(\mathbf{x},\mathbf{y}) = exp(-d(\mathbf{x},\mathbf{y}))$

Where are kernels used?

- **Nearest neighbors**
  - Measure similarity in the kernel space
- **Linear and logistic regression**
  - Map points to new, transformed feature space
- **SVMs and Perceptrons**
- **PCA**
  - SVD[X^TX]

What is the most common kernel method for linear regression?

What are we seeking to accomplish with kernels for classification?

What is the main benefit for PCA?

Is it Scale invariant?

- KNN
- Decision Trees
- Linear regression (OLS)
- Ridge regression
- Elastic net
- Logistic regression
- Kernel regression

What questions do you have on today's class?

Top

True or false

- Kernels in effect transform observations **x** to a higher dimension space **ϕ(x)**
  - **False**: It can be either higher or lower dimension
- Since kernels measure similarity,
  - k(x,y) < k(x,x) for x != y
  - **False.** If the kernel is derived from a distance metric (e.g. a Gaussian kernel), then that's true, but it is not true for e.g. the linear kernel
- If there exists a pair of points **x** and **y** such that **k(x,y) < 0**, then **k()** is not a kernel
  - **False**: kernels need to yield a positive semi-definite matrix, but individual entries in the matrix can be negative

True or false

- A quadratic kernel, when used in linear regression, gives results very similar to including quadratic interaction terms in the regression
  - **False:** when one includes quadratic interaction terms, that adds around $p^2/2$ new weights; the quadratic kernel does not introduce any new parameters.
- Any function **$\phi(x)$** can be used to generate a kernel using **$k(x,y) = \phi(x)^T\phi(y)$**
  - **True**
- Any distance metric **$d(x,y)$** can be used to generate a kernel using **$k(x,y) = \exp(-d(x,y))$**
  - **True**

Kernels form a dual representation

- Start with an ***n*p*** matrix ***X*** of predictors
- Generate an ***n*n*** *kernel matrix* ***K***
  - with elements $K_{ij} = k(x_i, x_j)$

We will cover and use this later!

Why is it not bad to generate a potentially much larger feature space?

The "kernel trick" avoids computing $\phi(x)$

- $k(x,y) = \phi(x)^T \phi(y)$
- So we can compute *k(x,y)* and never compute the expanded features $\phi(x)$

We will cover and use this later!

Gather.town

- https://gather.town/aQMGI0I1R8DP0Ovv/penn-cis