The Kardashian Kernel

David F. Fouhey

*Sublime and Distinguished Grand Poobah,*
CMU, Karlsruhe Inst. of Technology, Kharkiv Polytechnic Inst.

Daniel Maturana

*Distinguished Appointed Lecturer of Keeping it Real,*
CMU, KAIST, Kyushu Inst. of Technology
1. Introduction
   Motivation
   Related work

2. The Kardashian Kernel
   Formalities
   On Some Issues Raised by the Kardashian Kernel

3. Applications
   Kardashian SVM
   Graph Kardashiancian
   Kardashian Kopula

4. Conclusions and future work
• Kernel machines are popular
  • Have fancy math
  • They work well
• Kernel machines are popular
  • Have fancy math
  • They work well
• The Kardashians are popular
  • (TODO)
• Kernel machines are popular
  • Have fancy math
  • They work well
• The Kardashians are popular
  • (TODO)
• Why not combine them?
• Kronecker product
• Krylov subspace methods
• Kolmogorov axioms
• Kalman Filters
• Kent distribution
• Karhunen-Loève Transform
• Keypoint retrieval w/ K-d tree search
• Kriging (AKA Gaussian process regression)
• Kohonen maps (AKA Self-Organizing Maps)
• K-grams
• K-folds
• K-armed bandits
• ...
• Kronecker product
• Krylov subspace methods
• Kolmogorov axioms
• Kalman Filters
• Kent distribution
• Karhunen-Loève Transform
• Keypoint retrieval w/ K-d tree search
• Kriging (AKA Gaussian process regression)
• Kohonen maps (AKA Self-Organizing Maps)
• K-grams
• K-folds
• K-armed bandits
• ...
• Our approach: provably $k$-optimal, as our paper has significantly more $k$’s and substantially more pictures of the Kardashians
1 Introduction
   Motivation
   Related work

2 The Kardashian Kernel
   Formalities
   On Some Issues Raised by the Kardashian Kernel

3 Applications
   Kardashian SVM
   Graph Kardashianian
   Kardashian Kopula

4 Conclusions and future work
• Let $\mathcal{X}$ be an instance space.
• Let $\mathcal{X}$ be an instance space.
• The Kardashian Kernel is an inner product operator $K_K : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$.
Definitions

- Let $\mathcal{X}$ be an instance space.
- The Kardashian Kernel is an inner product operator $K_K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$.
- Kernel trick (Mercer): $K_K(x, x') = \kappa(x)^T \kappa(x)$, with $\kappa : \mathcal{X} \to \mathcal{K}$.
• Let $\mathcal{X}$ be an instance space.
• The Kardashian Kernel is an inner product operator $K_K : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$.
• Kernel trick (Mercer): $K_K(x, x') = \kappa(x)^T \kappa(x)$, with $\kappa : \mathcal{X} \rightarrow \mathbb{R}$.
• can leverage the Kardashian Feature space without suffering the Kurse of Dimensionality.
The Kardashian Kernel Trick

\[ \kappa : \mathbb{R}^n \rightarrow \text{Span}\left\{\right\} \]

\[ K_K(x, x') = \kappa(x)^T \kappa(x') \]
On Reproducing Kardashian Kernels

- Does $K_K$ define a Reproducing Kernel Hilbert Space (RKHS)? i.e. are the Kardashians Reproducing Kernels?
On Reproducing Kardashian Kernels

- Does $K_K$ define a Reproducing Kernel Hilbert Space (RKHS)? i.e. are the Kardashians Reproducing Kernels?
- Only proven for case of Kourtney
On Reproducing Kardashian Kernels

- Does $K_K$ define a Reproducing Kernel Hilbert Space (RKHS)? i.e. are the Kardashians Reproducing Kernels?
- Only proven for case of Kourtney
- But prominent bloggers argue that it is also true for Kim
On Divergence Functionals

Crucial question: does the space induced by $\kappa$ have structure that is advantageous to minimizing the $f$-divergences?

Theorem

$$\min_w = \frac{1}{n} \sum_{i=1}^{n} \langle w, \kappa(x_i) \rangle - \frac{1}{n} \sum_{j=1}^{n} \log \langle w, \kappa(y_j) \rangle + \frac{\lambda_n}{2} \| w \|^2_{K}$$

Proof.

Obvious by the use of the Jensen-Jenner Inequality.
1 Introduction
   Motivation
   Related work

2 The Kardashian Kernel
   Formalities
   On Some Issues Raised by the Kardashian Kernel

3 Applications
   Kardashian SVM
   Graph Kardashian
   Kardashian Kopula

4 Conclusions and future work
Regular Support Vector Machines (SVMs) are boring. We propose to solve the following optimization problem, which is subject to the Kardashian-Karush-Kuhn-Tucker (KKKT) Conditions:

$$\min_{w, \xi, b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{n} \xi_i$$
Kardashian SVM problem setting

Regular Support Vector Machines (SVMs) are boring. We propose to solve the following optimization problem, which is subject to the Kardashian-Karush-Kuhn-Tucker (KKKT) Conditions:

\[
\min_{\mathbf{w}, \xi, b} \frac{1}{2} \| \mathbf{w} \|^2 + C \sum_{i=1}^{n} \xi_i
\]

such that

\[
y_i (\mathbf{w}^T \kappa (\mathbf{x}_i) - b) \geq 1 - \xi_i \quad 1 \leq i \leq n
\]

\[
\xi_i \geq 0 \quad 1 \leq i \leq n
\]

\[
\zeta_j = 0 \quad 1 \leq j \leq m.
\]
- Standard approach: Kuadratic Programming (KP)
• Standard approach: Quadratic Programming (KP)
• But see Kurvature of optimization manifold
• Standard approach: Kuadratic Programming (KP)
• But see Kurvature of optimization manifold
• Take advantage of geometry: Konvex-Koncave Procedure (KKP)
Experiment: Kardashian or Cardassian?

Our “Kardashian or Cardassian” dataset.
In the feature space $\mathcal{F}$ induced by $\kappa$, the decision boundary between Cardassian and Kardashian lies approximately 5 light years from Cardassia Prime.
• The Graph Laplacian $\ell$

$$
\ell_{i,j} := \begin{cases} 
\deg(v_i) & \text{if } i = j \\
-1 & \text{if } i \neq j \text{ and } v_i \text{ is adjacent to } v_j \\
0 & \text{otherwise.}
\end{cases}
$$
• The Graph Kardashianian $\mathcal{K}$

\[ \mathcal{K}_{i,j} := \begin{cases} 
\deg(v_i) & \text{if } i = j \\
-\kappa & \text{if } i \neq j \text{ and } v_i \text{ is Kardashian-adjacent to } v_j \\
0 & \text{otherwise.}
\end{cases} \]
Applications

Graph Kardashianian

Graph Kardashianian

• Application: KardashianRank
• Application: KardashianRank
Applications

Kardashian Kopula

- Powerful generalization of the Gaussian Copula
- Video illustrating the Kardashian Kopula (featuring rapper Ray J) may be found in the supplementary material
• Powerful generalization of the Gaussian Copula

\[ c_\Sigma(u) = \frac{1}{\sqrt{\det \Sigma}} \exp \left( -\frac{1}{2} \Phi^{-1}(u)^T (\Sigma^{-1} - I) \Phi^{-1}(u) \right) \]
• Powerful generalization of the Gaussian Copula

\[ c^K_{\Sigma}(u) = \frac{1}{\sqrt{\text{det } \Sigma}} \exp \left( -\frac{1}{2} K^{-1}(u)^T (\Sigma^{-1} - I) K^{-1}(u) \right) \]
• Powerful generalization of the Gaussian Copula

\[ c_\Sigma^K(u) = \frac{1}{\sqrt{\text{det } \Sigma}} \exp \left( -\frac{1}{2} K^{-1}(u)^T (\Sigma^{-1} - I) K^{-1}(u) \right) \]

• Video illustrating the Kardashian Kopula (featuring rapper Ray J) may be found in the supplementary material
1. Introduction
   - Motivation
   - Related work

2. The Kardashian Kernel
   - Formalities
   - On Some Issues Raised by the Kardashian Kernel

3. Applications
   - Kardashian SVM
   - Graph Kardashianian
   - Kardashian Kopula

4. Conclusions and future work
We have exhausted Kardashianity, but currently working on:
Celebrity-based Machine Learning

The Tila Tequilla Transform ($T_{TT}$)

$T_{TT}(I)$
Celebrity-based Machine Learning

The Jensen-Shannon-Jersey-Shore ($JS^2$) divergence

$$JS_D(P \parallel Q) = \frac{1}{2} D(P \parallel M) + \frac{1}{2} D(Q \parallel M)$$

A powerful generalization of The Kardashian-Kulback-Leibler ($KKL$) divergence
Celebrity-based Machine Learning

Jamie Lee Curtis Regularization

$$\min_{\beta(t)} \left( \| y - \sum_{l=1}^{L} X_l \beta(t)_l \|^2_2 + \lambda \| \beta(t) - \beta(t-24h) \|^2_2 \right)$$
Celebrity-based Machine Learning

The Richard Pryor Prior
The Carrie Fisher Information Matrix

\[ \mathcal{I}(\theta) = E \left[ (\frac{\partial}{\partial \theta} \log f(x | \theta))^2 \right| \theta \]
Celebrity-based Machine Learning

Miley Cyrus Markov Chain Monte Carlo (MCMC) methods for inference
Celebrity-based Machine Learning

Hannah Montana Hidden Markov Models (HMHMHMM).

Train with MCMC for best of both worlds!
Celebrity-based Machine Learning

The Orlando Bloom Filter
Conclusions and future work

Celebrity-based Machine Learning

Johnny Depp Belief Nets (JDBNs)