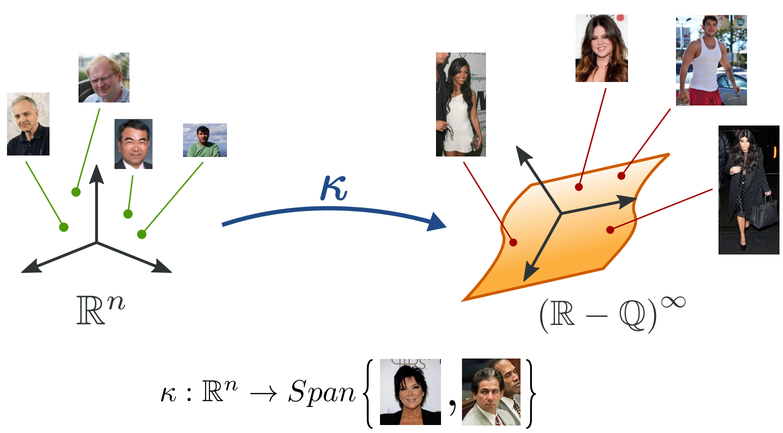
The Kardashian Kernel

David F. Fouhey
Sublime and Distinguished Grand Poobah,
Joint Appointment, Carnegie Mellon University,
Karlsruhe Inst. of Technology, Kharkiv Polytechnic Institute
dfouhey@cs.cmu.edu

Daniel Maturana
Distinguished Appointed Lecturer of Keeping it Real,
Joint Appointment, Carnegie Mellon University,
KAIST, Kyushu Inst. of Technology
dimatura@cmu.edu

Abstract

We propose a new family of kernels, based on the Kardashian family. We provide theoretical insights. We dance. As an application, we apply the new class of kernels to the problem of doing stuff.



$$K_K(x, x') = \kappa(x)^T \kappa(x')$$

Figure 1: A motivation for the Kernel Trick. κ maps the real world of sane people into the subset of $(\mathbb{R} - \mathbb{Q})^\infty$ spanned by Robert Kardashian, Sr. and Kris Jenner (formerly Kardashian). We would like to avoid having our data appear on Keeping Up with the Kardashians, and so we use the Kardashian Kernel K_K to compute an inner product in the Kardashian Space without ever having to go there.

Kernel machines are popular because they work well and have fancy math, pleasing practitioners and theoreticians alike. The Kardashians are popular because (**TODO**). Why not combine them?

1 The Kardashian Kernel

Let $\mathcal{X}$ be an instance space. The Kardashian Kernel is an inner product operator $K_K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. Applying the Kernel trick [14] we express it as $K_K(x, x') = \kappa(x)^T \kappa(x)$, with $\kappa : \mathcal{X} \to \mathfrak{K}$. Here $\mathfrak{K}$ represents a possibly infinitely-dimensional feature space. In Fig. 1, we provide the best (to our knowledge) motivation of the Kernel Trick: by using the Kardashian Kernel, we can leverage the Kardashian Feature space without suffering the Kurse of Dimensionality. This kurse is similar in nature to the better-known Curse of Dimensionality (c.f., [3]); however, the motivation for avoiding such a space is different: here, we wish to avoid having our data be associated with the Kardashian shenanigans[1].

1.1 Related Work

It is common in the literature to cite work that is thematically related; here, we explore an totally better style, in which we cite work that is alphabetically related.

Historically, we are of course motivated by the Kronecker product and delta, and by the fundamental impact had by Andrey Kolmogorov on probability theory. In more recent work, our work is motivated by Kalman Filters [9], especially in application to active stereo sensors such as Kinect or circular distributions such as the Kent distribution. We are also inspired by the ingenuity of David Lowe in using an modified K-d tree search ordering to perform fast keypoint retrieval in computer vision [11]; however, we believe that our approach is provably k-optimal, as our paper has significantly more k's and substantially more pictures of the Kardashians. We feel that it is important to note that several machine learning techniques are in fact, special cases of our k-themed approach: Gaussian process regression also known as Kriging [12], and Self-Organizing Maps [10] are also known as Kohonen maps, and thus both are of interest; in contrast to this work, however, we place our k's up-front systematically rather than hide them in complex formulations and semantics.

1.2 On Some Issues Raised by the Kardashian Kernel

1.2.1 On Reproducing Kardashian Kernels

A natural question is, does K_K define a Reproducing Kernel Hilbert Space (RKHS)? In other words, are the Kardashians Reproducing Kernels? We conjecture that it may be the case: see Fig. 2, as well as [16] and [13]. Nonetheless this has only been proven for a special case, Kourtney [5].

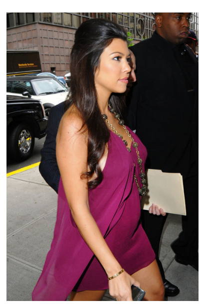


Figure 2: Are the Kardashians Reproducing Kernels? So far this conjecture has only been proven in the case of Kourtney (left figure), but many authors have argued that figures such as (right) may suggest it is also true for Kim.

[1]We thank the anonymous reviewer at *People* for correcting an error which occurred in the previous version: Kris Humphries is no longer considered a member of the Kardashian feature space.

1.2.2 On Divergence Functionals

Our paper naturally raises a number of questions. Most importantly of all, one must ask whether the space induced by κ has structure that is advantageous to minimizing the f-divergences (e.g., see [15])? We provide a brief analysis and proof sketch. Note that

$$D_\phi(\mathbb{Z}, \mathbb{Q}) = \int p_0 \phi(q_0/p_0) d\mu$$

with ϕ convex. The following result follows fairly straight-forwardly from the standard definitions:

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

A complete proof is omitted due to space considerations, but should be fairly straight-forward for even an advanced undergraduate; it is made much easier by the use of the Jensen-Jenner Inequality [8].

2 Kardashian SVM

2.1 Problem setting

SVMs are very popular, and provide a great way to plug in our new kernel and demonstrate the importance of being Kardashian [18]. We propose to solve the following optimization problem, which is subject to the Kardashian-Karush-Kuhn-Tucker (KKKT) Conditions[2]

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

such that

$$\begin{array}{ll} y_i(\mathbf{w}^T \kappa(\mathbf{x}_i) - \mathbf{b}) \geq 1 - \xi_i & 1 \leq i \leq n \\ \xi_i \geq 0 & 1 \leq i \leq n \\ \zeta_j = 0 & 1 \leq j \leq m. \end{array}$$

κ is the mapping of datapoints into the Kardashian feature space; $\mathbf{x}_i$ and $\mathbf{y}_i$ are data points $1, \ldots, n$ and their labels (-1 or 1); ξ_i are slack variables; and ζ_i are the sentences imposed upon O.J. Simpson, defended by Robert Kardashian, Sr., for charges $1, \ldots, m$. It can be proven that for $n = 3$, each ξ_i has the psychological interpretation of the expected relative proportion of attention given to Kardashian daughter i by the American public.

2.2 Learning algorithm

Learning the optimal classifier involves finding an optimal $\mathbf{w}$. A common approach is to use standard Kuadratic Programming (KP) methods; see [4] for an summary of relevant techniques[3].

However, the optimization manifold has a highly characteristic kurvature (see fig. 3). We use an alternative approach that takes advantage of the structure of the problem (c.f., our earlier discussion regarding minimal f-divergences in RKHS).

It is clear[4] that the problem meets the conditions to be considered "Konvex". Analogously, its dual is "Koncave". The solution for our problem is bounded by relaxed versions of both; therefore we use a Koncave-Konvex (KKP) procedure [19] to solve the problem.

2.3 Experiments - Kardashian or Cardassian?

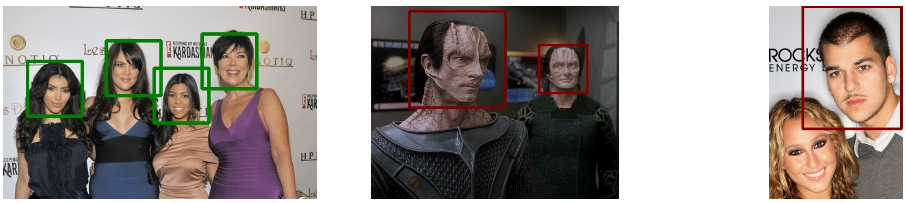
In this experiment, we use the proposed Kardashian-Support Vector Machine (K-SVM) to learn a classifier for "Cardassian" or "Kardashian" given a window of a humanoid face. The

[2] Conditions under which our function will converge to a global optimum and stay there for at least 72 days.

[3] N.B. Although popular, readers should note significant spelling deficiencies in [4]; notably, "konvex" is misspelled as "convex [sic]"

[4] Unless you're dumb.

Figure 3: Kurvature of the optimization manifold



(a) Kardashian - (l. to r.) Kim, Khloé, Kourtney, Kris

(b) Cardassian - (l. to r.) Gul Dukat, Elim Garak

(c) A failure case (or is it?): Kardashian - Rob

Figure 4: Example detections on our "Kardashian or Cardassian" dataset. After detection by a standard Humanoid Face detector, subjects are classified as Cardassian or Kardashian. Green bounding boxes indicate Kardashians; red ones indicate Cardassians.

Kardashians are a family of American socialites made famous by the purported reality show "Keeping up with the Kardashians"; the Cardassians are a species of humanoid extraterrestial beings from Cardassia prime, depicted in "Star Trek", a science fiction franchise.

We train and test on "Kardashian or Cardassian", a dataset which contains approximately 2000 examples of each class [5] Given a bounding box of a subject, we classify subjects into either Cardassian or Kardashian using HOG features. We demonstrate state-of-the art results, slightly out-performing a state-of-the-art latent SVM formulation in average precision, while beating it in Kardashianity[6]. It significantly beats both 1-NN and K-NN (not to be confused with Kardashian-Nearest Neighbors [1]) in performance as well. When paired with a standard humanoid facial detector, we can quickly detect and accurately localize and characterize persons of terrestial or interstellar origin as one of the two classes.

3 Kardashian Kernel Karhunen-Loève Transform

Just as the Kardashians have been used to deceptively market predatory credit cards [6], we use them to deceptively market an ineffective dimensionality reduction technique. Taking inspiration from their "famous-because-they're-famous" nature, we present a "low-dimension-because-it's-low dimension" dimensionality reduction technique, the Kardashian Kernel Karhunen-Loève Transform (KKKLT). Note that the Karhunen-Loève Transform (KLT) is also known as Principal Component Analysis (PCA), but we prefer the former nomenclature for its k-optimality. Our approach is highly general and works for arbitrary objective functions.

[5] Some may argue our ground truth is not be completely accurate, since there have been recent claims that Khloe is not a biological Kardashian [7]. However, we believe that this is not true and you should leave Khloe alone.

[6] Roughly equivalent to Rademacher complexity, measured in the Kardashian space $\mathfrak{K}$ induced by K_K, except for various constraints on the name of the algorithm and package used to implement it. We have yet to explore connections to Kolmogorov complexity.

Method	K-SVM	Latent SVM	K-NN	1-NN
AP	**80.1%**	78.1%	66.5%	60.9%
Kardashianity	∞	0	1	0

Table 1: Performance results on "Kardashian or Cardassian" dataset, with K-fold validation

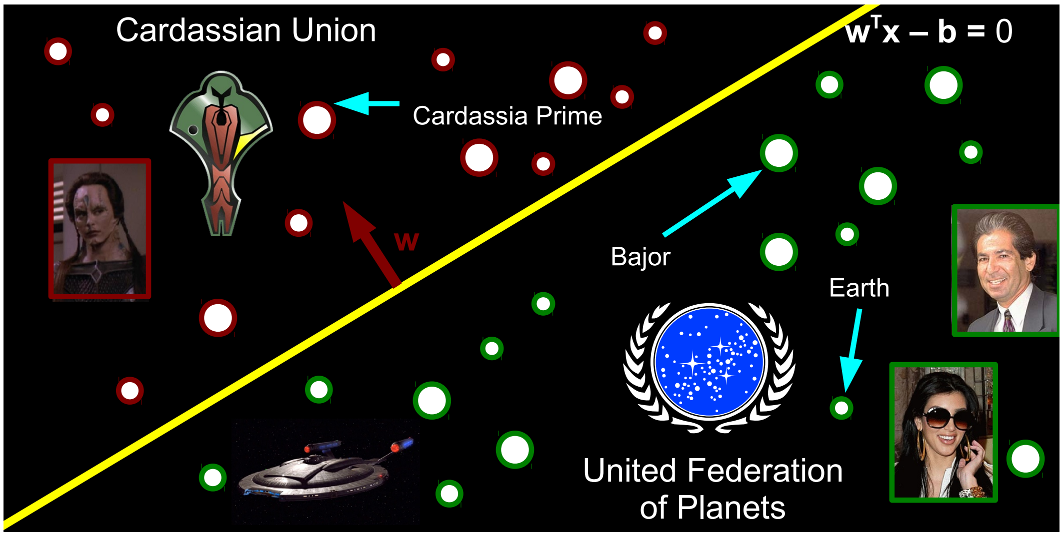


Figure 5: Schematic for Kardashian SVM: in the feature space $\mathfrak{K}$ induced by κ, the decision boundary between Cardassian and Kardashian lies approximately 5 light years from Cardassia Prime.

3.1 Method

Input: Datapoints $\mathbf{X} = \{\mathbf{x}_i : 1 \leq i \leq n, \mathbf{x}_i \in \mathfrak{R}^m\}$; objective function $F : \mathfrak{R}^{m \times n} \times \mathfrak{R}^{k \times n} \to \mathfrak{R}$; target dimension k; and k-dimensional representation $\mathbf{Y} = \{\mathbf{y}_i : 1 \leq i \leq n, \mathbf{y}_i \in \mathfrak{R}^k\}$ s.t. $F(\mathbf{X}, \mathbf{Y}) \leq F(\mathbf{X}, \mathbf{Y}') \; \forall \; \mathbf{Y}' \in \mathfrak{R}^{k \times n}$.

Output: k-dimensional representation of $\mathbf{X}$ (i.e., some $\mathbf{Z} \in \mathfrak{R}^{k \times n}$).

Procedure:

1. Compute the Kernelized Kovariance (see [17] for an illuminating discussion of the identical, but non-k-optimal, concept of Kernelized *Covariance*).
2. Compute the eigenvectors of the Kernelized Kovariance using the Kardashian-Krylov Subspace method; following k-optimality, this is preferred to Lanczos or Power Iteration.
3. Treat eigenvectors as a basis and form an infinite-dimensional rotation matrix R_R^R.
4. Compute the Kronecker product $R_R^R \otimes (R_R^R)^{-1}$, and take the SVD of the resulting matrix, again using Kardashian-Krylov methods for maximal k-optimality.
5. **return** $\mathbf{Y}$.

3.2 Experiments

Our approach is provably optimal for any objective function F (see Appendix A); no experiments are necessary. However, we use this space to pontificate, as is common in the literature. We believe that the Kardashian-inspired approach suggests underlying connections to the classic K-armed bandit problem. Further research will be needed. Please send cash (USD) to the authors to fund a reality show/NSF REU on this.

4 Graph Kardashiancian

In Kardashian Kernelized space the weight of data points is defined by the weight of the data points to which they are linked to. In this situation the graph connectivity is essential in the ranking of celebrities; e.g., see Fig. 6, which explains the prominence of Kim Kardashian[7] To capture this connectivity we define an analog of the so-called Laplacian on graph similarity matrices, the Kardashiancian. For details on the Kardanshiancian, and its applications, see our forthcoming paper on our Kardashiancian-powered algorithm, KardashianRank [2].

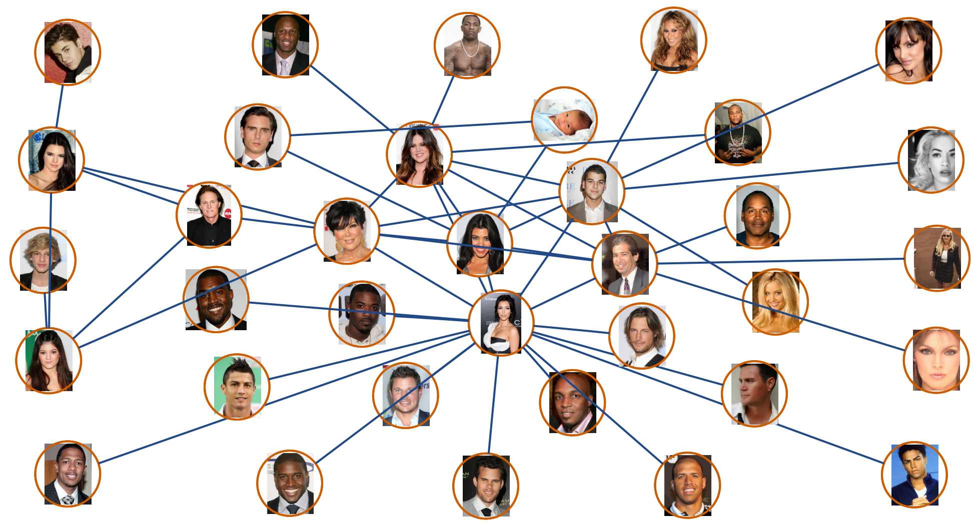


Figure 6: KardashianRank: Kardashiancian-based celebrity ranking

5 Conclusions and future work

We consider the Kardashian Kernel the foundational breakthrough of a new field, Celebrity-based Machine Learning (CML). We believe that we have already exhausted the potential for Kardashianity in this, and other K-Themed papers (e.g., [1, 2]); however, while we have closed the door on any new Kardashian papers by our thorough and complete treatment of the topic, various avenues of new research are now wide open - we recommend young researchers to hop on the bandwagon as soon as possible. We are currently working on various new groundbreaking results:

- The Tila Tequila Transform (T_{T_T})
- Johnny Depp Belief Nets ($JDBN$s)
- The Kardashian-Kulback-Leibler (KKL) divergence and its generalization, the Jensen-Shannon-Jersey-Shore (JS^2) divergence
- Miley Cyrus Markov Chain Monte Carlo ($MCMCMC$) methods for inference
- The Orlando Bloom Filter
- The Carrie Fisher Information Matrix

Our work suggests the monetization of research via branding: for instance, in exchange for a significant monetary compensation, we suggest that the NIPS and ICML insist upon the convention of referring to "Tikhonov regularization" as "Jamie Lee Curtis Regularization"®, sponsored by Activia Yogurt™.

[7]Please see supplementary video for an insightful and intuitive explanation of the connection between Kim Kardashian and Ray J.

Acknowledgments

This paper is brought to you by the letter K, the restaurants on Kraig Street, and contributions from viewers like you.

A Proof that Kardashian Kernel KLT/PCA is optimal

Suppose $\mathbf{Z} \in \Re^{k \times n}$ is the output of KKKLT$(\mathbf{X}, \mathbf{Y}, k, F)$; then $\mathbf{Z}$ satisfies $F(\mathbf{X}, \mathbf{Z}) \leq F(\mathbf{X}, \mathbf{Z}') \; \forall \; \mathbf{Z}' \in \Re^{k \times n}$.

Proof. Trivially follows from the input conditions. $\square$

References

[1] Anonymous. Kardashian Nearest Neighbors - toward a totally new era of non-parametric methods. *In Submission to Transactions on Pattern Analysis and Machine Intelligence*, 2012.

[2] Anonymous. KardashianRank: A random STD model for celebrity ranking. *In Submission to Transactions on Computer Networks and ISDN Systems*, 2012.

[3] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, 1st ed. 2006. corr. 2nd printing edition, 2007.

[4] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[5] E!Online. *The Internet*, 2012. URL: `http://www.eonline.com/news/kourtney_kardashian_pregnant_with_baby/277501`.

[6] Gawker. *The Internet*, 2010. URL: `http://gawker.com/5693964/kim-kardashians-credit-card-may-be-the-worst-credit-card-ever`.

[7] Perez Hilton. *The Internet*, 2012. URL: `http://perezhilton.com/2012-01-11-khloe-kardashian-reportedly-not-a-biological-kardashian`.

[8] K. Jenner and R. Kardashian. The Jensen-Jenner inequality. *IEEE Transations on Los Angeles-Themed Statistics*, 13(6):1352–1368, mar. 1998.

[9] R.E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45.

[10] Teuvo Kohonen. *Self-Organizing Maps*. Springer, 2001.

[11] David G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 2004.

[12] G. Matheron. Principles of geostatistics. *Economic Geology*, (58):1246–1266, 1963.

[13] MediaTakeout. *The Internet*, 2010. URL: `http://mediatakeout.com/45282/mto_world_exclusive_kim_kardashian_is_reportedly_pregnant____and_youll_never_guess_who_the_father_is.html`.

[14] John Mercer. Functions of positive and negative type and their connection with the theory of integral equations. *Philos. Trans. Roy. Soc. London*, 1909.

[15] XuanLong Nguyen, M.J. Wainwright, and M.I. Jordan. Estimating divergence functionals and the likelihood ratio by convex risk minimization. *Information Theory, IEEE Transactions on*, 56(11):5847 –5861, nov. 2010.

[16] RadarOnline. *The Internet*, 2012. URL: `http://www.radaronline.com/exclusives/2012/01/khloe-kardashian-fertility-treatments-lamar-odom-dallas`.

[17] L. Song, A. Gretton, D. Bickson, Y. Low, and C. Guestrin. Kernel belief propagation. 2011.

[18] O. Wilde. *The Importance of Being Kardashian*. Random House, 1895.

[19] A.L. Yuille, A. Rangarajan, K. Kardashian, K. Kardashian, and K. Kardashian. The koncave-konvex procedure. *Neural Komputation*, 15(4):915–936, 2003.