Kernels

Zoltán Szabó – CMAP

B.Sc. Day @ École Polytechnique

September 4, 2018

Inner product $\rightarrow$ kernel: similarity between features

Extension of $k(x,y) = \langle x, y \rangle = \sum_i x_i y_i$:

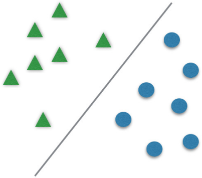
$$k(x,y) = \langle \varphi(x), \varphi(y) \rangle_{\mathcal{H}}.$$

Inner product → kernel: similarity between features

Extension of $k(x,y) = \langle x, y \rangle = \sum_i x_i y_i$:

$$k(x,y) = \langle \varphi(x), \varphi(y) \rangle_{\mathcal{H}}.$$

- Classification (SVM):

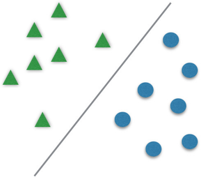


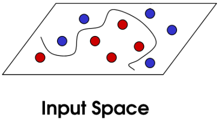
Inner product $\rightarrow$ kernel: similarity between features

Extension of $k(x,y) = \langle x, y \rangle = \sum_i x_i y_i$:

$$k(x,y) = \langle \varphi(x), \varphi(y) \rangle_{\mathcal{H}}.$$

- Classification (SVM):

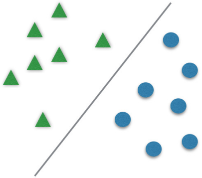


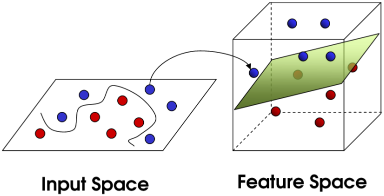


Extension of $k(x,y) = \langle x, y \rangle = \sum_i x_i y_i$:

$$k(x,y) = \langle \varphi(x), \varphi(y) \rangle_{\mathcal{H}}.$$

- Classification (SVM):



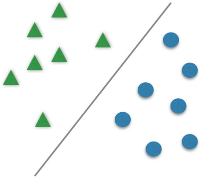


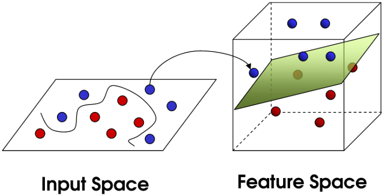
Inner product → kernel: similarity between features

Extension of $k(x, y) = \langle x, y \rangle = \sum_i x_i y_i$:

$$k(x, y) = \langle \varphi(x), \varphi(y) \rangle_{\mathcal{H}}.$$

- Classification (SVM):

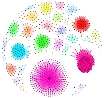




- Representation of distributions: $\mathbb{P} \mapsto \mathbb{E}_{x \sim \mathbb{P}} \varphi(x)$.

Kernels: examples





- $\mathcal{X} = \mathbb{R}^d$, $\gamma > 0$, $p \in \mathbb{Z}^+$:

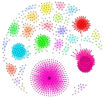
$$k(x, y) = \langle \varphi(x), \varphi(y) \rangle_{\mathcal{H}}$$

$$k_p(x, y) = (\langle x, y \rangle + \gamma)^p, \qquad k_G(x, y) = e^{-\gamma \|x-y\|_2^2},$$

$$k_e(x, y) = e^{-\gamma \|x-y\|_2}, \qquad k_C(x, y) = 1 + \frac{1}{\gamma \|x - y\|_2^2}.$$

Kernels: examples





- $\mathcal{X} = \mathbb{R}^d$, $\gamma > 0$, $p \in \mathbb{Z}^+$: $k(x,y) = \langle \varphi(x), \varphi(y) \rangle_{\mathcal{H}}$

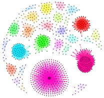
$$k_p(x,y) = (\langle x,y \rangle + \gamma)^p, \qquad k_G(x,y) = e^{-\gamma \|x-y\|_2^2},$$

$$k_e(x,y) = e^{-\gamma \|x-y\|_2}, \qquad k_C(x,y) = 1 + \frac{1}{\gamma \|x-y\|_2^2}.$$

- $\mathcal{X}$ = strings (DNA, ...):
  - r-spectrum kernel: # of common $\leqslant r$-substrings.

Kernels: examples





- $\mathcal{X} = \mathbb{R}^d$, $\gamma > 0$, $p \in \mathbb{Z}^+$:

$$k(x,y) = \langle \varphi(x), \varphi(y) \rangle_{\mathcal{H}}$$

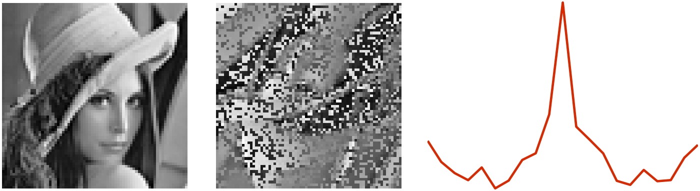
$$k_p(x,y) = (\langle x,y \rangle + \gamma)^p, \qquad k_G(x,y) = e^{-\gamma \|x-y\|_2^2},$$

$$k_e(x,y) = e^{-\gamma \|x-y\|_2}, \qquad k_C(x,y) = 1 + \frac{1}{\gamma \|x-y\|_2^2}.$$

- $\mathcal{X}$ = strings (DNA, ...):
  - r-spectrum kernel: # of common $\leqslant r$-substrings.
- $\mathcal{X}$ = time series: dynamic time-warping.

Outlier-robust image registration

Given two images:

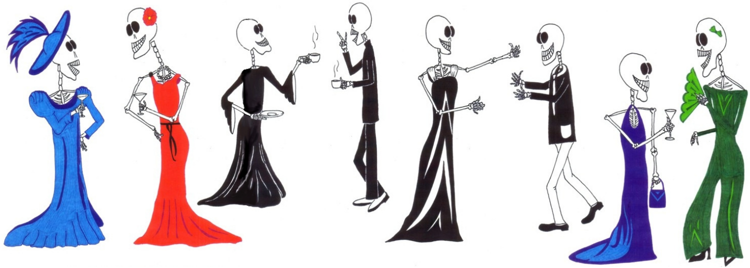


Goal: find the transformation which takes the right one to the left.

Independent subspace analysis

Cocktail party problem:

- independent groups of people / music bands,
- observation = mixed sources.

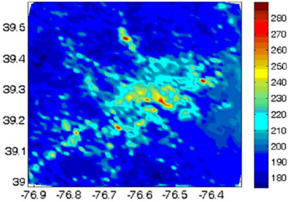


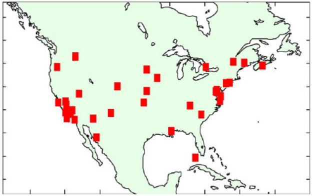
Goal: demixing.

Distribution regression: aerosol prediction

- **Goal**: aerosol prediction = air pollution → climate.

- Prediction using labelled bags:
  - bag := multi-spectral satellite measurements over an area,
  - label := local aerosol value.



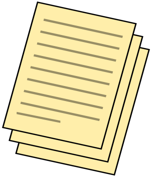


Feature selection

- **Goal**: find
  - the feature subset (# of rooms, criminal rate, local taxes)
  - most relevant for house price prediction (y).

2-sample testing: natural language processing

- Given: 2 categories of documents.
- Example: Bayesian inference, neuroscience.
- **Task**:
  - test their distinguishability,
  - most discriminative words $\rightarrow$ interpretability.

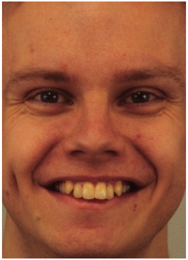


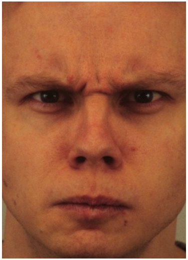
2-sample testing: natural language processing

- Given: 2 categories of documents.
- Example: Bayesian inference, neuroscience.
- **Task**:
  - test their distinguishability,
  - most discriminative words → interpretability.

Do $\{x_i\}$ and $\{y_j\}$ come from the same distribution?

2-sample testing: computer vision





- Given: two sets of faces (happy, angry).
- **Task**:
  - check if they are different,
  - determine the most discriminative features/regions.

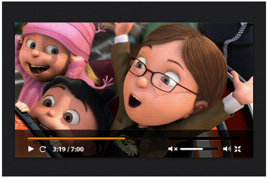
Independence testing-1: media annotations

- We are given paired samples. **Task**: test independence.
- Examples:
  - (song, year of release) pairs

Independence testing-1: media annotations

- We are given paired samples. **Task**: test independence.
- Examples:
  - (song, year of release) pairs

  - (video, caption) pairs



Independence testing-2: translation

- How do we detect dependency? (paired samples)

x_1: Honourable senators, I have a question for the Leader of the Government in the Senate with regard to the support funding to farmers that has been announced. Most farmers have not received any money yet.

x_2: No doubt there is great pressure on provincial and municipal governments in relation to the issue of child care, but the reality is that there have been no cuts to child care funding from the federal government to the provinces. In fact, we have increased federal investments for early childhood development.

...

y_1: Honorables sénateurs, ma question s'adresse au leader du gouvernement au Sénat et concerne l'aide financiére qu'on a annoncée pour les agriculteurs. La plupart des agriculteurs n'ont encore rien reu de cet argent.

y_2: Il est évident que les ordres de gouvernements provinciaux et municipaux subissent de fortes pressions en ce qui concerne les services de garde, mais le gouvernement n'a pas réduit le financement qu'il verse aux provinces pour les services de garde. Au contraire, nous avons augmenté le financement fédéral pour le développement des jeunes enfants.

...

Independence testing-2: translation

- How do we detect dependency? (paired samples)

x_1: Honourable senators, I have a question for the Leader of the Government in the Senate with regard to the support funding to farmers that has been announced. Most farmers have not received any money yet.

x_2: No doubt there is great pressure on provincial and municipal governments in relation to the issue of child care, but the reality is that there have been no cuts to child care funding from the federal government to the provinces. In fact, we have increased federal investments for early childhood development.

...

y_1: Honorables sénateurs, ma question s'adresse au leader du gouvernement au Sénat et concerne l'aide financiére qu'on a annoncée pour les agriculteurs. La plupart des agriculteurs n'ont encore rien reu de cet argent.

y_2: Il est évident que les ordres de gouvernements provinciaux et municipaux subissent de fortes pressions en ce qui concerne les services de garde, mais le gouvernement n'a pas réduit le financement qu'il verse aux provinces pour les services de garde. Au contraire, nous avons augmenté le financement fédéral pour le développement des jeunes enfants.

...

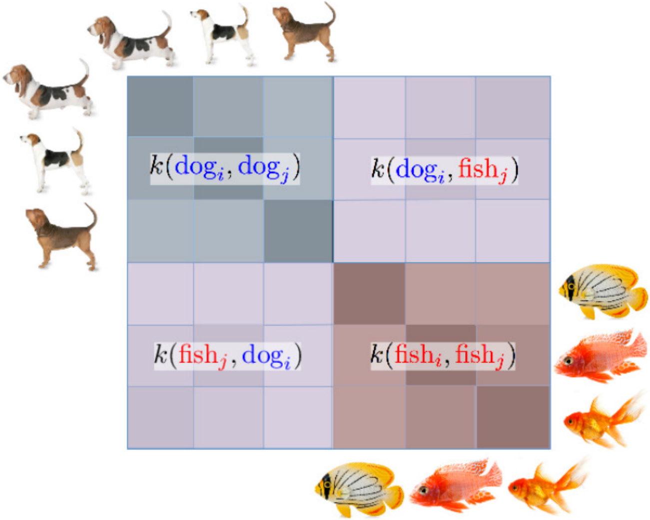
Are the French paragraphs translations of the English ones, or have nothing to do with them?

MMD estimator: intuition

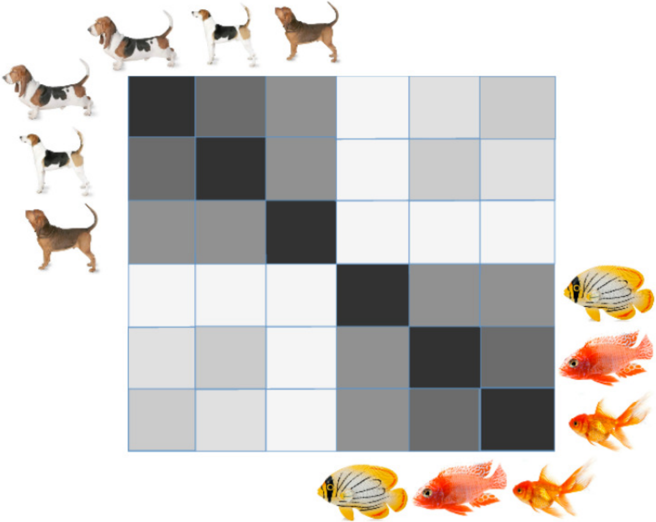
$$\sim P$$

$$\sim Q$$

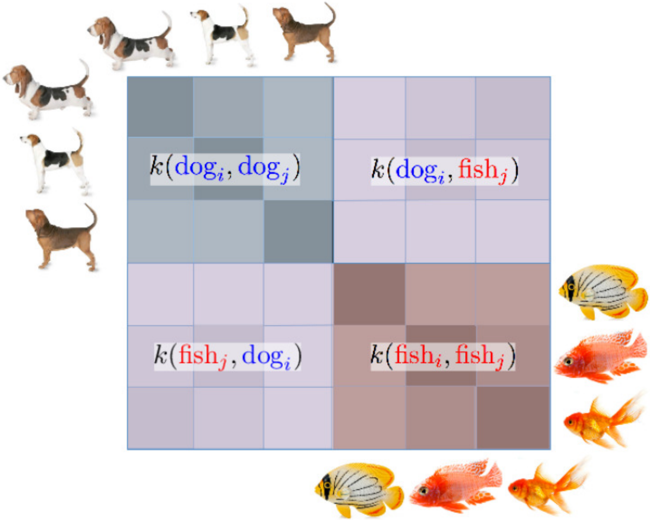
MMD estimator: intuition



MMD estimator: intuition



MMD estimator: intuition



$\widehat{\mathrm{MMD}}^2(\mathbb{P},\mathbb{Q}) = \overline{G_{\mathbb{P},\mathbb{P}}} + \overline{G_{\mathbb{Q},\mathbb{Q}}} - 2\overline{G_{\mathbb{P},\mathbb{Q}}}$ (without diagonals in $\overline{G_{\mathbb{P},\mathbb{P}}}$, $\overline{G_{\mathbb{Q},\mathbb{Q}}}$)

† $\widehat{\mathrm{MMD}}$ & $\widehat{\mathrm{HSIC}}$ illustration credit: Arthur Gretton

HSIC: intuition. $\mathcal{X}$: images, $\mathcal{Y}$: descriptions

Their noses guide them through life, and they're never happier than when following an interesting scent. They need plenty of exercise, about an hour a day if possible.

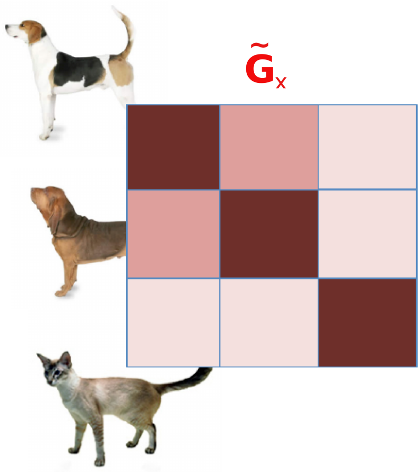
A large animal who slings slobber, exudes a distinctive houndy odor, and wants nothing more than to follow his nose. They need a significant amount of exercise and mental stimulation.

Known for their curiosity, intelligence, and excellent communication skills, the Javanese breed is perfect if you want a responsive, interactive pet, one that will blow in your ear and follow you everywhere.

Text from dogtime.com and petfinder.com

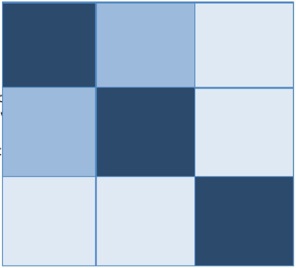
HSIC intuition: Gram matrices

$\tilde{G}_x$



Their noses guide them through life, and they're never happier than when following an interesting scent. They need plenty of exercise, about an hour a day if possible.

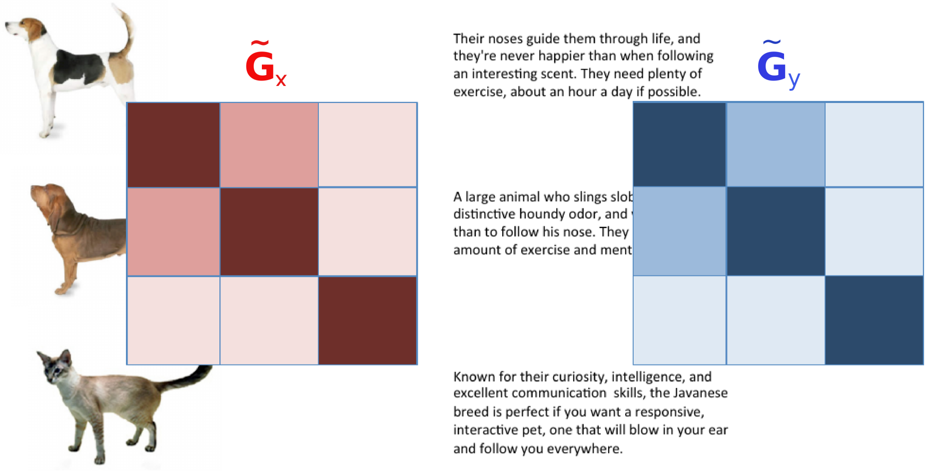
$\tilde{G}_y$



A large animal who slings slob
distinctive houndy odor, and
than to follow his nose. They
amount of exercise and ment

Known for their curiosity, intelligence, and excellent communication skills, the Javanese breed is perfect if you want a responsive, interactive pet, one that will blow in your ear and follow you everywhere.

HSIC intuition: Gram matrices



Empirical estimate:

$$\widehat{\mathrm{HSIC}^2}(\mathbb{P}_{xy}) = \frac{1}{n^2}\left\langle \tilde{\mathbf{G}}_x, \tilde{\mathbf{G}}_y \right\rangle_F.$$

Links

ITE, webpage:

- https://bitbucket.org/szzoli/ite-in-python/
- http://www.cmap.polytechnique.fr/~zoltan.szabo/

Links

ITE, webpage:

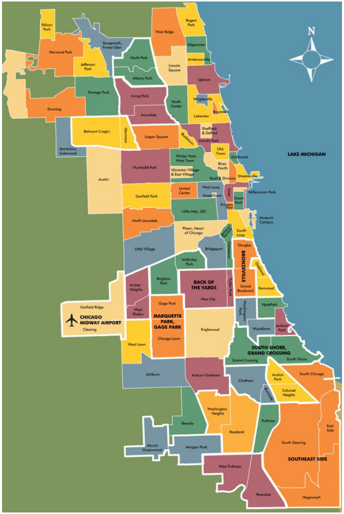
- https://bitbucket.org/szzoli/ite-in-python/
- http://www.cmap.polytechnique.fr/∼zoltan.szabo/

Thank you for the attention!

Some kernel-enriched domains: $\varphi \rightarrow k$

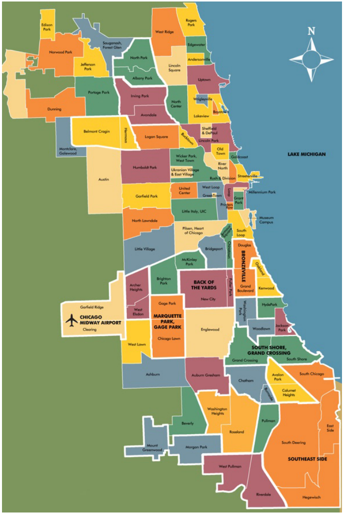
- Trees [Collins and Duffy, 2001, Kashima and Koyanagi, 2002], time series [Cuturi, 2011], strings [Lodhi et al., 2002],
- mixture models, hidden Markov models or linear dynamical systems [Jebara et al., 2004],
- sets [Haussler, 1999, Gärtner et al., 2002], fuzzy domains [Guevara et al., 2017], distributions [Hein and Bousquet, 2005, Martins et al., 2009, Muandet et al., 2011],
- groups [Cuturi et al., 2005] $\xrightarrow{\text{spec.}}$ permutations [Jiao and Vert, 2016],
- graphs [Vishwanathan et al., 2010, Kondor and Pan, 2016].

- Given:
  - Density/model: p.



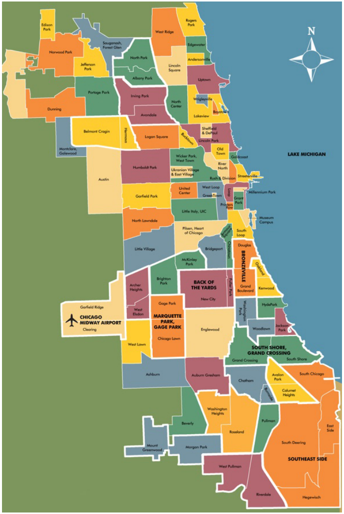
Goodness-of-fit testing: criminal data analysis

- Given:
  - Density/model: p.
  - Samples: $X = \{x_i\}_{i=1}^n \sim q$ (unknown).



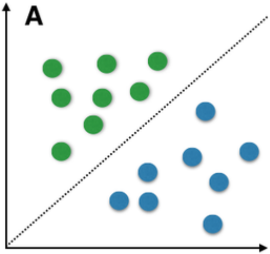
- Given:
  - Density/model: p.
  - Samples: $X = \{x_i\}_{i=1}^n \sim q$ (unknown).
- Task: using p, X test

$$H_0 : p = q, \quad \text{vs}$$
$$H_1 : p \neq q.$$



Classification motivation: linear separability

Idealized situation

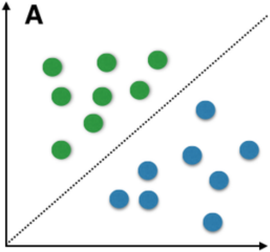


Decision surface:

$$\{\mathbf{x} : \langle \mathbf{w}, \mathbf{x} \rangle = 0\}$$

Classification motivation: linear separability

Idealized situation



Decision surface:

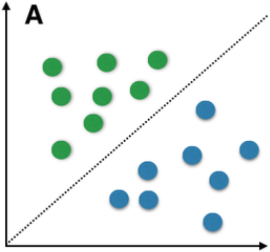
$$\{\mathbf{x} : \langle \mathbf{w}, \mathbf{x} \rangle = 0\} \Rightarrow$$

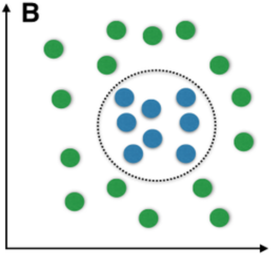
classes:

$$\{\mathbf{x} : \langle \mathbf{w}, \mathbf{x} \rangle \geqslant 0\} \qquad \{\mathbf{x} : \langle \mathbf{w}, \mathbf{x} \rangle < 0\}$$

Classification motivation: non-linear separability

Idealized situation | Real world





Decision surface (left):

$$\{\mathbf{x} : \langle \mathbf{w}, \mathbf{x} \rangle = 0\} \Rightarrow$$

classes:

$$\{\mathbf{x} : \langle \mathbf{w}, \mathbf{x} \rangle \geqslant 0\} \qquad \{\mathbf{x} : \langle \mathbf{w}, \mathbf{x} \rangle < 0\}.$$

On the ellipse

$$\left\{\mathbf{x} : \frac{(x_1 - c_1)^2}{a^2} + \frac{(x_2 - c_2)^2}{b^2} = 1\right\}$$

Non-linear separability – continued

On the ellipse, outside

$$\left\{\mathbf{x} : \frac{(x_1 - c_1)^2}{a^2} + \frac{(x_2 - c_2)^2}{b^2} = 1\right\},$$

$$\left\{\mathbf{x} : \frac{(x_1 - c_1)^2}{a^2} + \frac{(x_2 - c_2)^2}{b^2} > 1\right\}$$

On the ellipse, outside, inside:

$$\left\{\mathbf{x} : \frac{(x_1 - c_1)^2}{a^2} + \frac{(x_2 - c_2)^2}{b^2} = 1\right\},$$

$$\left\{\mathbf{x} : \frac{(x_1 - c_1)^2}{a^2} + \frac{(x_2 - c_2)^2}{b^2} > 1\right\},$$

$$\left\{\mathbf{x} : \frac{(x_1 - c_1)^2}{a^2} + \frac{(x_2 - c_2)^2}{b^2} < 1\right\}.$$

Non-linear separability – continued

On the ellipse, outside, inside:

$$\left\{\mathbf{x} : \frac{(x_1 - c_1)^2}{a^2} + \frac{(x_2 - c_2)^2}{b^2} = 1\right\},$$

$$\left\{\mathbf{x} : \frac{(x_1 - c_1)^2}{a^2} + \frac{(x_2 - c_2)^2}{b^2} > 1\right\},$$

$$\left\{\mathbf{x} : \frac{(x_1 - c_1)^2}{a^2} + \frac{(x_2 - c_2)^2}{b^2} < 1\right\}.$$

With polynomial feature: $\varphi(\mathbf{x}) = (x_1^2, x_1, 1, x_2^2, x_2)$:

- Decision surface: $\{\mathbf{x} : \langle \mathbf{w}, \varphi(\mathbf{x}) \rangle = 0\}$.

Non-linear separability – continued

On the ellipse, outside, inside:

$$\left\{\mathbf{x} : \frac{(x_1 - c_1)^2}{a^2} + \frac{(x_2 - c_2)^2}{b^2} = 1\right\},$$

$$\left\{\mathbf{x} : \frac{(x_1 - c_1)^2}{a^2} + \frac{(x_2 - c_2)^2}{b^2} > 1\right\},$$

$$\left\{\mathbf{x} : \frac{(x_1 - c_1)^2}{a^2} + \frac{(x_2 - c_2)^2}{b^2} < 1\right\}.$$

With polynomial feature: $\varphi(\mathbf{x}) = (x_1^2, x_1, 1, x_2^2, x_2)$:

- Decision surface: $\{\mathbf{x} : \langle \mathbf{w}, \varphi(\mathbf{x}) \rangle = 0\}$.
- Classes: $\{\mathbf{x} : \langle \mathbf{w}, \varphi(\mathbf{x}) \rangle > 0\}$, $\{\mathbf{x} : \langle \mathbf{w}, \varphi(\mathbf{x}) \rangle < 0\}$.

Quadratic & polynomial features

Still in $\mathbb{R}^2$:

$$\varphi(\mathbf{x}) = \left(x_1^2, \sqrt{2}x_1x_2, x_2^2\right),$$

$$\langle \varphi(\mathbf{x}), \varphi(\mathbf{x}') \rangle = ?$$

Quadratic & polynomial features

Still in $\mathbb{R}^2$:

$$\varphi(\mathbf{x}) = \left(x_1^2, \sqrt{2}x_1x_2, x_2^2\right),$$

$$\langle\varphi(\mathbf{x}), \varphi(\mathbf{x}')\rangle = \left\langle \begin{bmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \end{bmatrix}, \begin{bmatrix} (x_1')^2 \\ \sqrt{2}(x_1')(x_2') \\ (x_2')^2 \end{bmatrix} \right\rangle$$

Quadratic & polynomial features

Still in $\mathbb{R}^2$:

$$\varphi(\mathbf{x}) = \left(x_1^2, \sqrt{2}x_1x_2, x_2^2\right),$$

$$\langle\varphi(\mathbf{x}), \varphi(\mathbf{x}')\rangle = \left\langle \begin{bmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \end{bmatrix}, \begin{bmatrix} (x_1')^2 \\ \sqrt{2}(x_1')(x_2') \\ (x_2')^2 \end{bmatrix} \right\rangle$$

$$= x_1^2(x_1')^2 + \underbrace{\sqrt{2}\sqrt{2}}_{2}\, x_1x_2(x_1')(x_2') + x_2^2(x_2')^2$$

Quadratic & polynomial features

Still in $\mathbb{R}^2$:

$$\varphi(\mathbf{x}) = \left(x_1^2, \sqrt{2}x_1x_2, x_2^2\right),$$

$$\langle \varphi(\mathbf{x}), \varphi(\mathbf{x}') \rangle = \left\langle \begin{bmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \end{bmatrix}, \begin{bmatrix} (x_1')^2 \\ \sqrt{2}(x_1')(x_2') \\ (x_2')^2 \end{bmatrix} \right\rangle$$

$$= x_1^2(x_1')^2 + \underbrace{\sqrt{2}\sqrt{2}}_{2} x_1x_2(x_1')(x_2') + x_2^2(x_2')^2$$

$$= \left(x_1x_1' + x_2x_2'\right)^2$$

Quadratic & polynomial features

Still in $\mathbb{R}^2$:

$$\begin{aligned}
\varphi(\mathbf{x}) &= \left(x_1^2, \sqrt{2}x_1x_2, x_2^2\right), \\
\langle \varphi(\mathbf{x}), \varphi(\mathbf{x}') \rangle &= \left\langle \begin{bmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \end{bmatrix}, \begin{bmatrix} (x_1')^2 \\ \sqrt{2}(x_1')(x_2') \\ (x_2')^2 \end{bmatrix} \right\rangle \\
&= x_1^2(x_1')^2 + \underbrace{\sqrt{2}\sqrt{2}}_{2} x_1x_2(x_1')(x_2') + x_2^2(x_2')^2 \\
&= \left(x_1x_1' + x_2x_2'\right)^2 \\
&= \left\langle \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \begin{bmatrix} x_1' \\ x_2' \end{bmatrix} \right\rangle^2 = \langle \mathbf{x}, \mathbf{x}' \rangle^2 =: k(\mathbf{x}, \mathbf{x}').
\end{aligned}$$

Quadratic & polynomial features

Still in $\mathbb{R}^2$:

$$\begin{aligned}
\varphi(\mathbf{x}) &= \left(x_1^2, \sqrt{2}x_1x_2, x_2^2\right), \\
\langle \varphi(\mathbf{x}), \varphi(\mathbf{x}') \rangle &= \left\langle \begin{bmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \end{bmatrix}, \begin{bmatrix} (x_1')^2 \\ \sqrt{2}(x_1')(x_2') \\ (x_2')^2 \end{bmatrix} \right\rangle \\
&= x_1^2(x_1')^2 + \underbrace{\sqrt{2}\sqrt{2}}_{2} x_1x_2(x_1')(x_2') + x_2^2(x_2')^2 \\
&= \left(x_1x_1' + x_2x_2'\right)^2 \\
&= \left\langle \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \begin{bmatrix} x_1' \\ x_2' \end{bmatrix} \right\rangle^2 = \langle \mathbf{x}, \mathbf{x}' \rangle^2 =: k(\mathbf{x}, \mathbf{x}').
\end{aligned}$$

$\langle \mathbf{x}, \mathbf{x}' \rangle^d = \langle \varphi(\mathbf{x}), \varphi(\mathbf{x}') \rangle$: $\varphi(\mathbf{x}) = d$-order polynomial. $\Rightarrow$

Quadratic & polynomial features

Still in $\mathbb{R}^2$:

$$\varphi(\mathbf{x}) = \left(x_1^2, \sqrt{2}x_1x_2, x_2^2\right),$$

$$\begin{aligned}
\langle\varphi(\mathbf{x}), \varphi(\mathbf{x}')\rangle &= \left\langle \begin{bmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \end{bmatrix}, \begin{bmatrix} (x_1')^2 \\ \sqrt{2}(x_1')(x_2') \\ (x_2')^2 \end{bmatrix} \right\rangle \\
&= x_1^2(x_1')^2 + \underbrace{\sqrt{2}\sqrt{2}}_{2} x_1x_2(x_1')(x_2') + x_2^2(x_2')^2 \\
&= \left(x_1x_1' + x_2x_2'\right)^2 \\
&= \left\langle \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \begin{bmatrix} x_1' \\ x_2' \end{bmatrix} \right\rangle^2 = \langle\mathbf{x}, \mathbf{x}'\rangle^2 =: k(\mathbf{x}, \mathbf{x}').
\end{aligned}$$

$\langle\mathbf{x}, \mathbf{x}'\rangle^d = \langle\varphi(\mathbf{x}), \varphi(\mathbf{x}')\rangle$: $\varphi(\mathbf{x}) = d$-order polynomial. $\Rightarrow$ Explicit computation would be heavy!

Collins, M. and Duffy, N. (2001).
Convolution kernels for natural language.
In *Advances in Neural Information Processing Systems (NIPS)*, pages 625–632.

Cuturi, M. (2011).
Fast global alignment kernels.
In *International Conference on Machine Learning (ICML)*, pages 929–936.

Cuturi, M., Fukumizu, K., and Vert, J.-P. (2005).
Semigroup kernels on measures.
Journal of Machine Learning Research, 6:1169–1198.

Gärtner, T., Flach, P. A., Kowalczyk, A., and Smola, A. (2002).
Multi-instance kernels.
In *International Conference on Machine Learning (ICML)*, pages 179–186.

Guevara, J., Hirata, R., and Canu, S. (2017).
Cross product kernels for fuzzy set similarity.
In *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–6.

Haussler, D. (1999).
Convolution kernels on discrete structures.
Technical report, Department of Computer Science, University of California at Santa Cruz.
(http://cbse.soe.ucsc.edu/sites/default/files/convolutions.pdf).

Hein, M. and Bousquet, O. (2005).
Hilbertian metrics and positive definite kernels on probability measures.
In *International Conference on AI and Statistics (AISTATS)*, pages 136–143.

Jebara, T., Kondor, R., and Howard, A. (2004).
Probability product kernels.

Journal of Machine Learning Research, 5:819–844.

Jiao, Y. and Vert, J.-P. (2016).
The Kendall and Mallows kernels for permutations.
In *International Conference on Machine Learning (PMLR)*, volume 37, pages 2982–2990.

Kashima, H. and Koyanagi, T. (2002).
Kernels for semi-structured data.
In *International Conference on Machine Learning (ICML)*, pages 291–298.

Kondor, R. and Pan, H. (2016).
The multiscale Laplacian graph kernel.
In *Advances in Neural Information Processing Systems (NIPS)*, pages 2982–2990.

Lodhi, H., Saunders, C., Shawe-Taylor, J., Cristianini, N., and Watkins, C. (2002).
Text classification using string kernels.
Journal of Machine Learning Research, 2:419–444.

Martins, A. F. T., Smith, N. A., Xing, E. P., Aguiar, P. M. Q., and Figueiredo, M. A. T. (2009).
Nonextensive information theoretic kernels on measures.
The Journal of Machine Learning Research, 10:935–975.

Muandet, K., Fukumizu, K., Dinuzzo, F., and Schölkopf, B. (2011).
Learning from distributions via support measure machines.
In *Neural Information Processing Systems (NIPS)*, pages 10–18.

Vishwanathan, S. N., Schraudolph, N. N., Kondor, R., and Borgwardt, K. M. (2010).
Graph kernels.
Journal of Machine Learning Research, 11:1201–1242.