On Canonical Correlation Analysis

CS189/289A: Introduction to Machine Learning

Stella Yu

UC Berkeley

25 September 2018
27 September 2018

Outline

1. Why CCA
2. Pearson Correlation Coefficient
3. Canonical Correlation Analysis
4. Understanding CCA
5. CCA Regression
6. Point Set Examples

Scale Issues in Learning

1. Feature collection: different types of measurements
2. Feature engineering: derived features, e.g. polynomials
3. Feature scaling: different units / weights of measurements

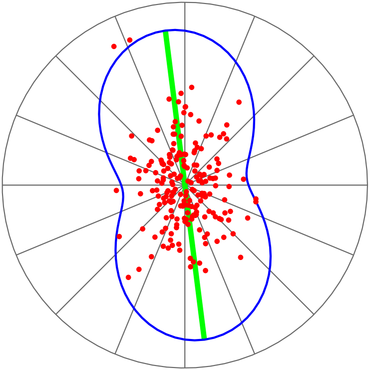
Common Solutions to Scale Issues

1. Feature collection: different types of measurements
   Feature standardization ?

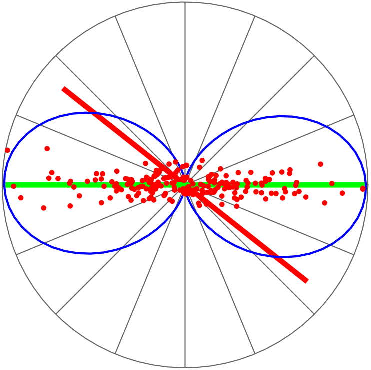
2. Feature engineering: derived features, e.g. polynomials
   Data matrix conditioning ?

3. Feature scaling: different units / weights of measurements
   ?

Scale Impact on Learning: PCA



(x, y)



$(2.5x, 0.25y)$

Why Canonical Correlation Analysis?

- Given two sets of random variables, there are correlations among the variables, CCA finds linear combinations of each set which have maximum correlation with each other.

- Max correlation $\neq$ Max variance.

- CCA: linear dependence on a common latent space H

$$X_{n\times p} + N_x = H_{n\times k}U_{k\times p} + A \cdot N_A \tag{1}$$

$$Y_{n\times q} + N_y = H_{n\times k}V_{k\times q} + B \cdot N_B \tag{2}$$

- Assume paired data (X, Y).

- Cross-correlation is invariant to scaling.

Covariance and Pearson's Correlation Coefficient

- Covariance between two random variables:

$$\mathrm{cov}(X,Y) = E\left[\, (X - E[X]) \cdot (Y - E[Y]) \,\right] \tag{3}$$

$$\mathrm{cov}(X,X) = E\left[\, (X - E[X])^2 \,\right] = V[X] \tag{4}$$

$$\mathrm{cov}(X,Y) = 0, \text{ if } X \text{ and } Y \text{ are independent} \tag{5}$$

- Population Pearson's correlation coefficient:

$$\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{cov}(X,X) \cdot \mathrm{cov}(Y,Y)}} = \frac{\mathrm{cov}(X,Y)}{\sqrt{V[X] \cdot V[Y]}} \tag{6}$$

$$\rho(Y,X) = \rho(X,Y) \tag{7}$$

$$-1 \leq \rho(X,Y) \leq 1 \tag{8}$$

- $\rho(X,Y)$ is not defined, when $V[X] = 0$ or $V[Y] = 0$.

Linear vs. Nonlinear Correlation vs. Independence

- Pearson's correlation detects only linear dependencies:

$$Y - E[Y] = k \cdot (X - E[X]) \tag{9}$$

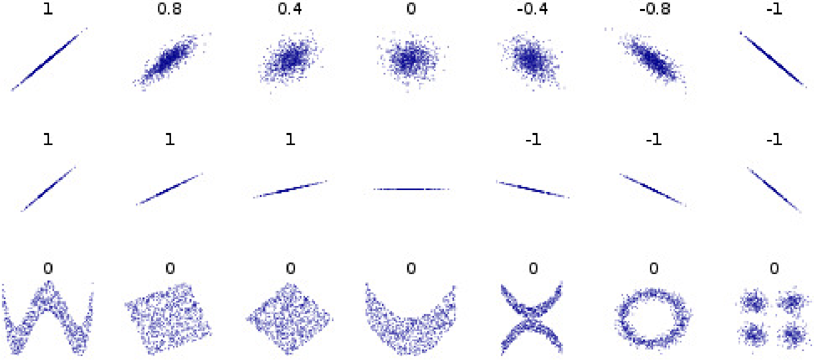
$$\Rightarrow \quad \rho(X,Y) = \frac{kV[X]}{\sqrt{V[X] \cdot k^2 V[X]}} = \pm 1, \quad \forall k. \tag{10}$$

- If X and Y are independent, then $\rho(X,Y) = 0$.

- If $\rho(X,Y) = 0$, then X and Y are linearly uncorrelated. They can be nonlinearly correlated and perfectly dependent, e.g. $Y = X^2, E[X] = 0$.

- If $\rho(X,Y) = 0$, when X and Y are jointly normal, uncorrelatedness is equivalent to independence.

Sample Correlation Coefficient: $\rho \to r$



$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}, \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad (11)$$

$$r_{xy} = \frac{x'y}{\sqrt{x'x \cdot y'y}}, \qquad x \Leftarrow x - \bar{x}, \quad y \Leftarrow y - \bar{y} \quad (12)$$

Key: Correlation Coefficient Is Affine Invariant

$$
\begin{aligned}
\rho(aX+c, bY+d) &= \rho(aX, bY) && (13)\\
&= \frac{\mathrm{cov}(aX, bY)}{\sqrt{V[aX]\cdot V[bY]}} && (14)\\
&= \frac{a\cdot b\cdot \mathrm{cov}(X, Y)}{\sqrt{a^2\cdot b^2\cdot V[X]\cdot V[Y]}} && (15)\\
&= \frac{\mathrm{cov}(X, Y)}{\sqrt{V[X]\cdot V[Y]}} && (16)\\
&= \rho(X, Y) && (17)
\end{aligned}
$$

Gaussian Distribution and Correlation Coefficient

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}(0, \Sigma) \tag{18}$$

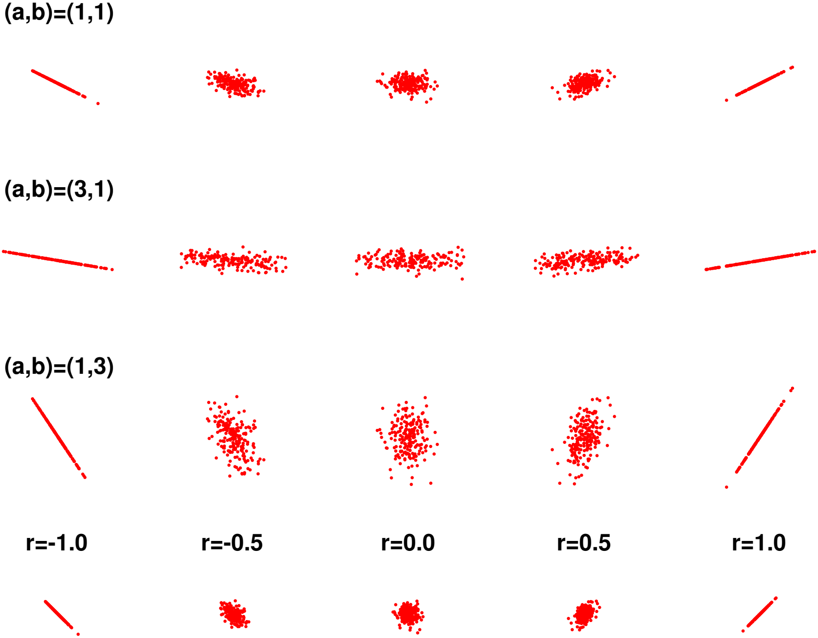
$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \tag{19}$$

$$\Sigma = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix} = \begin{bmatrix} \sigma_X^2 & \rho \cdot \sigma_X \sigma_Y \\ \rho \cdot \sigma_X \sigma_Y & \sigma_Y^2 \end{bmatrix} \tag{20}$$

$$\begin{bmatrix} \sigma_X^{-1} & 0 \\ 0 & \sigma_Y^{-1} \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} \sigma_X^{-1} & 0 \\ 0 & \sigma_Y^{-1} \end{bmatrix} \Sigma \begin{bmatrix} \sigma_X^{-1} & 0 \\ 0 & \sigma_Y^{-1} \end{bmatrix}'\right) \tag{21}$$

$$\sim \mathcal{N}\left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right) \tag{22}$$

Gaussian Distribution and Correlation Coefficient



Canonical Correlation Analysis (CCA)

- Paired data matrices $(X_{n\times p}, Y_{n\times q})$, zero-mean
  n paired points in p and q dimensional spaces respectively.

- Simultaneously find projection directions $u_{p\times 1}$ in the X space and $v_{q\times 1}$ in the Y space such that the projected data onto u and v have maximal correlation:

$$\max_{u,v} \varepsilon(u,v;X,Y) = \rho(Xu, Yv) \tag{23}$$

$$= \frac{\mathrm{cov}(Xu, Yv)}{\sqrt{V[Xu]\cdot V[Yv]}} = \frac{u'X'Yv}{\sqrt{u'X'Xu \cdot v'Y'Yv}} \tag{24}$$

- In general, CCA seek a latent basis dimension k, $k \leq \min(p,q)$, where total projection correlations are maximized.

CCA Invariant to Scaling or Affine Transform

- Suppose (u^*, v^*) are the optimal projections for (X, Y):

$$(u^*, v^*) = \arg\max_{u,v} \rho(Xu, Yv) = \frac{u'X'Yv}{\sqrt{u'X'Xu \cdot v'Y'Yv}} \tag{25}$$

- If X and Y are linearly transformed:

$$\tilde{X} = XA \tag{26}$$
$$\tilde{Y} = YB \tag{27}$$

- The optimal correlation value does not change:

$$\max_{u,v} \rho(\tilde{X}u, \tilde{Y}v) = \rho(XAA^{-1}u^*, YBB^{-1}v^*) = \max_{u,v} \rho(Xu, Yv) \tag{28}$$

- The optimal projections are transformed accordingly:

$$(A^{-1}u^*, B^{-1}v^*) = \arg\max_{u,v} \rho(\tilde{X}u, \tilde{Y}v) \tag{29}$$

CCA Criterion and Its Lagrangian

- CCA criterion as a normalized correlation ratio:

$$\max_{u,v} \varepsilon(u,v;X,Y) = \frac{u'X'Yv}{\sqrt{u'X'Xu \cdot v'Y'Yv}} \tag{30}$$

- Reformulate it into a constrained optimization problem:

$$\begin{aligned}
&\max_{u,v} && u'C_{xy}v && (31)\\
&\text{s. t.} && u'C_{xx}u = 1, \quad v'C_{yy}v = 1 && (32)\\
&\text{where} && C_{xy} = X'Y && (33)\\
& && C_{xx} = X'X && (34)\\
& && C_{yy} = Y'Y && (35)
\end{aligned}$$

- Lagrangian:

$$L(u,v;a,b) = u'C_{xy}v - a(u'C_{xx}u - 1) - b(v'C_{yy}v - 1) \tag{36}$$

CCA Solution: Optimality Conditions

- Lagrangian:

$$L(u, v; a, b) = u' C_{xy} v - a \cdot (u' C_{xx} u - 1) - b \cdot (v' C_{yy} v - 1) \quad (37)$$

- Optimality conditions:

$$L_u = C_{xy} v - a \cdot C_{xx} u = 0 \quad \Rightarrow \quad C_{xy} v = a \cdot C_{xx} u \quad (38)$$

$$L_v = C'_{xy} u - b \cdot C_{yy} v = 0 \quad \Rightarrow \quad C'_{xy} u = b \cdot C_{yy} v \quad (39)$$

$$L_a = u' C_{xx} u - 1 = 0 \quad \Rightarrow \quad u' C_{xx} u = 1 \quad (40)$$

$$L_b = v' C_{yy} v - 1 = 0 \quad \Rightarrow \quad v' C_{yy} v = 1 \quad (41)$$

- Solving the dual variables:

$$u' C_{xy} v = a \cdot u' C_{xx} u = a \quad (42)$$

$$v' C'_{xy} u = b \cdot v' C_{yy} v = b \quad (43)$$

$$a = b = \lambda \quad (44)$$

Eigensolution to CCA

- Coupled equations:

$$C_{xy}v = \lambda \cdot C_{xx}u \tag{45}$$

$$C_{yx}u = \lambda \cdot C_{yy}v \tag{46}$$

$$C_{yx} = C'_{xy} \tag{47}$$

- Eigensolution in the joint project direction space:

$$\begin{bmatrix} & C_{xy} \\ C_{yx} & \end{bmatrix}\begin{bmatrix} u \\ v \end{bmatrix} = \lambda \begin{bmatrix} C_{xx} & \\ & C_{yy} \end{bmatrix}\begin{bmatrix} u \\ v \end{bmatrix} \tag{48}$$

- Optimum:

$$\max_{u,v} \quad \varepsilon(u, v; X, Y) = \lambda_1 \tag{49}$$

$$\begin{bmatrix} u \\ v \end{bmatrix} = \mathrm{eig}_1\left(\begin{bmatrix} & C_{xy} \\ C_{yx} & \end{bmatrix}, \begin{bmatrix} C_{xx} & \\ & C_{yy} \end{bmatrix}\right) \tag{50}$$

Solution Relations to Linear Subspace Methods

PCA ,PLS (partial least squares), MLR (multivariate linear regression), and CCA share the same generalized eigensolution for Rayleigh quotient optimization:

$$\max_{w} \qquad \varepsilon(w; M, D) = \frac{w'Mw}{w'Dw} \tag{51}$$

$$\text{optimum:} \qquad Mw = \lambda Dw \tag{52}$$

method	M	D
PCA / TLS	C_{xx}	I
PLS	$\begin{bmatrix} 0 & C_{xy} \\ C_{yx} & 0 \end{bmatrix}$	$\begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$
MLS	$\begin{bmatrix} 0 & C_{xy} \\ C_{yx} & 0 \end{bmatrix}$	$\begin{bmatrix} C_{xx} & 0 \\ 0 & I \end{bmatrix}$
CCA	$\begin{bmatrix} 0 & C_{xy} \\ C_{yx} & 0 \end{bmatrix}$	$\begin{bmatrix} C_{xx} & 0 \\ 0 & C_{yy} \end{bmatrix}$

SVD: Duality and Eigen in the Joint Space

$$(U_i, V_i, s_i) = \mathrm{SVD}(X) \tag{53}$$

$$X = USV' \tag{54}$$

$$XV_i = s_i U_i \tag{55}$$

$$X'U_i = s_i V_i \tag{56}$$

$$\begin{bmatrix} & X \\ X' & \end{bmatrix} \begin{bmatrix} U_i \\ V_i \end{bmatrix} = s_i \begin{bmatrix} U_i \\ V_i \end{bmatrix} \tag{57}$$

$$\left(\begin{bmatrix} U_i \\ V_i \end{bmatrix}, s_i \right) = \mathrm{eig}\left(\begin{bmatrix} & X \\ X' & \end{bmatrix} \right) \tag{58}$$

SVD: Align Projections According to A Metric

$$\max_{u,v} \quad \text{align}(u,v;M) = u'Mv \tag{59}$$

$$\text{s. t.} \quad \|u\| = 1, \quad \|v\| = 1 \tag{60}$$

$$M = USV' \tag{61}$$

$$u = U\alpha, v = V\beta, \quad \text{s. t.} \quad \|\alpha\| = 1, \|\beta\| = 1 \tag{62}$$

$$\text{align}(u,v;M) = u'Mv = \alpha' U' M V \beta \tag{63}$$

$$= \alpha' S \beta \tag{64}$$

$$\leq \left(\frac{S\beta}{\|S\beta\|} \right)' (S\beta) \tag{65}$$

$$= \|S\beta\| = \sqrt{\beta_1^2 s_1^2 + \beta_2^2 s_2^2 + \dots} \tag{66}$$

$$\leq s_1 \tag{67}$$

$$\beta = \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix}', \alpha = \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix}' \tag{68}$$

$$u = U\alpha = U_1 \tag{69}$$

$$v = V\beta = V_1 \tag{70}$$

SVD View of CCA Solution

$$\begin{bmatrix} & C_{xy} \\ C_{yx} & \end{bmatrix}\begin{bmatrix} u \\ v \end{bmatrix} = \lambda \begin{bmatrix} C_{xx} & \\ & C_{yy} \end{bmatrix}\begin{bmatrix} u \\ v \end{bmatrix} \tag{71}$$

$$\begin{bmatrix} C_{xx} & \\ & C_{yy} \end{bmatrix}^{-\frac{1}{2}}\begin{bmatrix} & C_{xy} \\ C_{yx} & \end{bmatrix}\begin{bmatrix} C_{xx} & \\ & C_{yy} \end{bmatrix}^{-\frac{1}{2}} \cdot \begin{bmatrix} C_{xx} & \\ & C_{yy} \end{bmatrix}^{\frac{1}{2}}\begin{bmatrix} u \\ v \end{bmatrix}$$

$$= \lambda \begin{bmatrix} C_{xx} & \\ & C_{yy} \end{bmatrix}^{\frac{1}{2}}\begin{bmatrix} u \\ v \end{bmatrix} \tag{72}$$

$$\begin{bmatrix} \tilde{u} \\ \tilde{v} \end{bmatrix} = \begin{bmatrix} C_{xx} & \\ & C_{yy} \end{bmatrix}^{\frac{1}{2}}\begin{bmatrix} u \\ v \end{bmatrix} \tag{73}$$

$$\begin{bmatrix} & C_{xx}^{-\frac{1}{2}} C_{xy} C_{yy}^{-\frac{1}{2}} \\ C_{yy}^{-\frac{1}{2}} C_{yx} C_{xx}^{-\frac{1}{2}} & \end{bmatrix}\begin{bmatrix} \tilde{u} \\ \tilde{v} \end{bmatrix} = \lambda \begin{bmatrix} \tilde{u} \\ \tilde{v} \end{bmatrix} \tag{74}$$

$$(\tilde{u}, \tilde{v}, \lambda) = \mathrm{svd}(C_{xx}^{-\frac{1}{2}} C_{xy} C_{yy}^{-\frac{1}{2}}) \tag{75}$$

Understanding CCA Step 1: Whitening

$$\begin{bmatrix} \tilde{u} \\ \tilde{v} \end{bmatrix} = \begin{bmatrix} C_{xx} & \\ & C_{yy} \end{bmatrix}^{\frac{1}{2}} \begin{bmatrix} u \\ v \end{bmatrix} \quad \Leftarrow \text{change of basis} \tag{76}$$

$$\begin{bmatrix} \tilde{X} \\ \tilde{Y} \end{bmatrix} = \begin{bmatrix} X \\ Y \end{bmatrix} \begin{bmatrix} C_{xx} & \\ & C_{yy} \end{bmatrix}^{-\frac{1}{2}} \quad \Leftarrow \text{change of coordinates} \tag{77}$$

$$\rho(Xu, Yv) = \rho(XC_{xx}^{-\frac{1}{2}} \cdot C_{xx}^{\frac{1}{2}} u, YC_{yy}^{-\frac{1}{2}} \cdot C_{yy}^{\frac{1}{2}} v) = \rho(\tilde{X}\tilde{u}, \tilde{Y}\tilde{v}) \tag{78}$$

$$= \frac{u' C_{xy} v}{\sqrt{u' C_{xx} u \cdot v' C_{yy} v}} = \frac{\tilde{u}' C_{xx}^{-\frac{1}{2}} C_{xy} C_{yy}^{-\frac{1}{2}} \tilde{v}}{\sqrt{\tilde{u}'\tilde{u} \cdot \tilde{v}'\tilde{v}}} = \frac{\tilde{u}' C_{\tilde{x}\tilde{y}} \tilde{v}}{\sqrt{\tilde{u}'\tilde{u} \cdot \tilde{v}'\tilde{v}}} \tag{79}$$

$$C_{\tilde{x}\tilde{y}} = \tilde{X}'\tilde{Y} = C_{xx}^{-\frac{1}{2}} C_{xy} C_{yy}^{-\frac{1}{2}} \tag{80}$$

$$C_{\tilde{x}\tilde{x}} = \tilde{X}'\tilde{X} = (XC_{xx}^{-\frac{1}{2}})' XC_{xx}^{-\frac{1}{2}} = C_{xx}^{-\frac{1}{2}} C_{xx} C_{xx}^{-\frac{1}{2}} = I \tag{81}$$

$$C_{\tilde{y}\tilde{y}} = \tilde{Y}'\tilde{Y} = I \tag{82}$$

That is, $\tilde{X}$ and $\tilde{Y}$ are whitened in their own spaces: decorrelated and isotropic along all the directions.

Understanding CCA Step 2: Align by Correlation

$$
\begin{aligned}
\max_{u,v} \quad \rho(Xu, Yv) &= \max_{\tilde{u},\tilde{v}} \rho(\tilde{X}\tilde{u}, \tilde{Y}\tilde{v}) && (83)\\
&= \frac{\tilde{u}' C_{\tilde{x}\tilde{y}} \tilde{v}}{\sqrt{\tilde{u}'\tilde{u} \cdot \tilde{v}'\tilde{v}}} && (84)\\
(\tilde{u}, \tilde{v}, \lambda) &= \mathrm{svd}(C_{\tilde{x}\tilde{y}}) && (85)\\
&= \mathrm{svd}(C_{xx}^{-\frac{1}{2}} C_{xy} C_{yy}^{-\frac{1}{2}}) && (86)
\end{aligned}
$$

Note: k-CCA are orthogonal in the whitened feature spaces.

Understanding CCA Step 3: Backprojection

$$(\tilde{u}, \tilde{v}, \lambda) = \mathrm{svd}(C_{xx}^{-\frac{1}{2}} C_{xy} C_{yy}^{-\frac{1}{2}}) \tag{87}$$

$$C_{xx}^{-\frac{1}{2}} C_{xy} C_{yy}^{-\frac{1}{2}} = \lambda \tilde{u} \tilde{v}' \tag{88}$$

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} C_{xx} & \\ & C_{yy} \end{bmatrix}^{-\frac{1}{2}} \begin{bmatrix} \tilde{u} \\ \tilde{v} \end{bmatrix} \tag{89}$$

Note: Unlike PCA, k-CCA are not orthogonal in the original spaces.

CCA: Too Good To Be True?

- For paired data (X, Y), we compute the CCA with the optimal correlation value λ.
- Does a large λ mean significant correlation?
  No, it could come from the small denominator of the ratio:

$$\rho(Xu, Yv) = \frac{u' C_{xy} v}{\sqrt{u' C_{xx} u \cdot v' C_{yy} v}} \tag{90}$$

- How to tell if the large correlation value arises by accident?
  Change the pairing of (X, Y) without changing the covariance of X or Y. If the correlation remains high (statistical test for significance on Wilks' λ), then it is due to the small covariances in the denominator, not to the large correlation between paired data in the numerator.

CCA and Projection Regression: Training

1. Given zero-mean training data (X, Y), compute CCA (U, V):

$$(U_{p\times k}, V_{q\times k}) = \mathrm{CCA}(X, Y) \tag{91}$$

2. Given (X, Y, U, V), compute their individual CCA projections:

$$X_c = X_{n\times p} U_{p\times k} \tag{92}$$

$$Y_c = Y_{n\times q} V_{q\times k} \tag{93}$$

   Note that (X_c, Y_c) is also of zero mean.

3. Given (X_c, Y_c), fit a $k \times k$ linear regressor A

$$Y_c = X_c A_{k\times k} \tag{94}$$

Prediction from CCA Projections: Testing

- Given CCA basis (U, V) and coefficient regressor A from the training data, given test data X, predict projection Y_c:

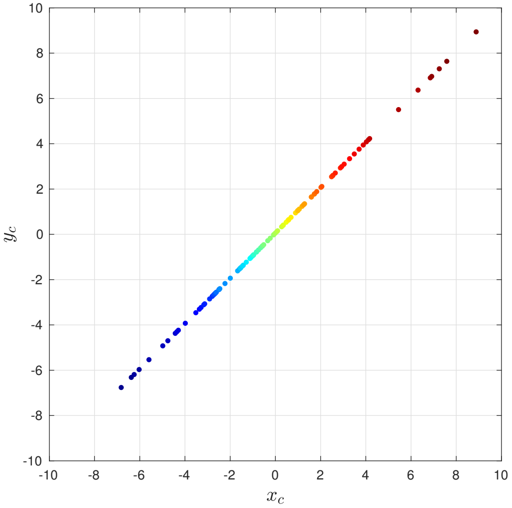
$$X_c = XU \tag{95}$$

$$\hat{Y}_c = X_c A = XUA \tag{96}$$

- Predict Y from projection Y_c:

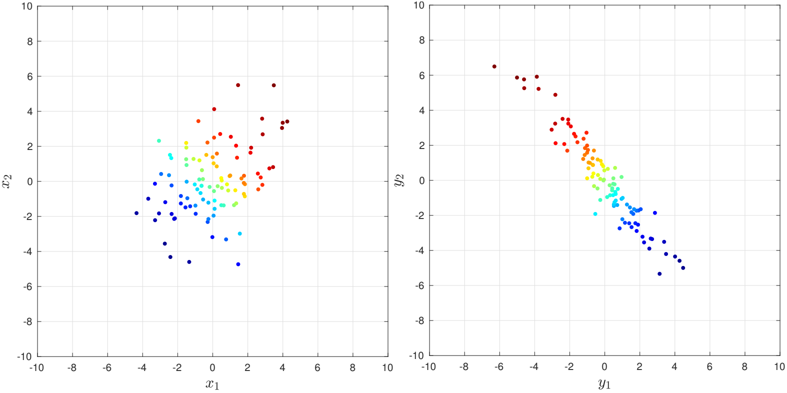
$$\hat{Y} = \hat{Y}_c (V'V)^{-1} V' \tag{97}$$

Point Set #1: Hidden Correlation Between Spaces



$$y_c = f(x_c) = k \cdot x_c \tag{98}$$

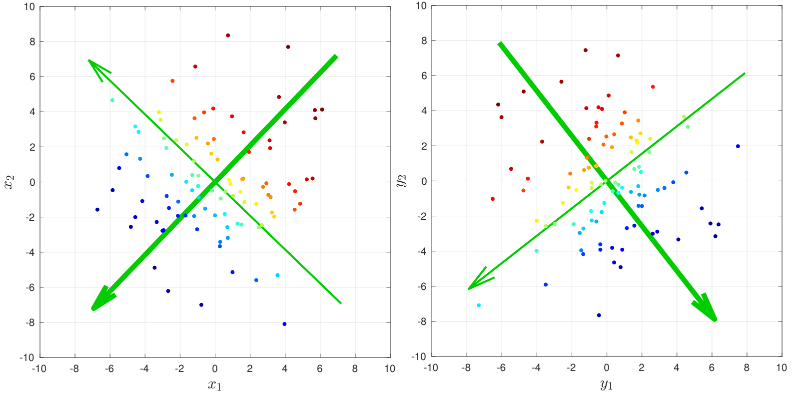
Irrelevant Orthogonal Components in Each Space



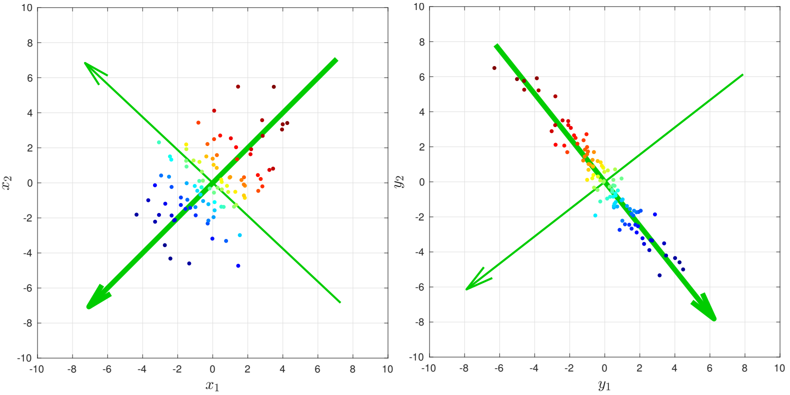
$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \cos(\alpha) & \sin(\alpha) \\ -\sin(\alpha) & \cos(\alpha) \end{bmatrix} \begin{bmatrix} x_c \\ x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \cos(\beta) & \sin(\beta) \\ -\sin(\beta) & \cos(\beta) \end{bmatrix} \begin{bmatrix} y_c \\ y_n \end{bmatrix}$$

$$x_n \sim \mathcal{N}(0, \sigma_x^2) \qquad y_n \sim \mathcal{N}(0, \sigma_y^2)$$

CCA in the Whitened Spaces: Orthogonal



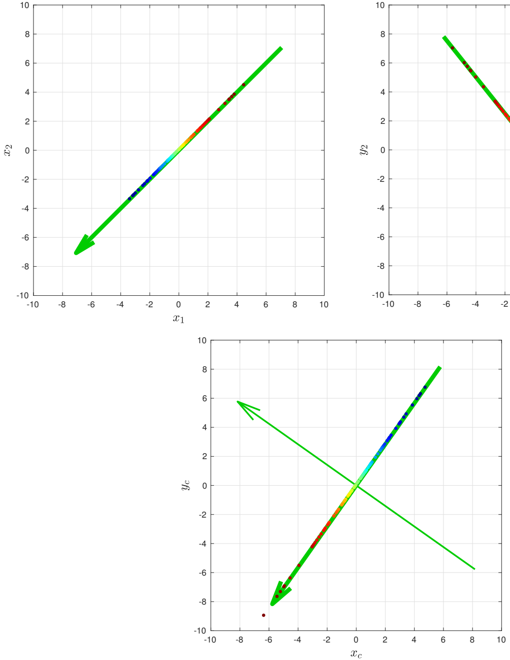
CCA in the Original Spaces: Orthogonal



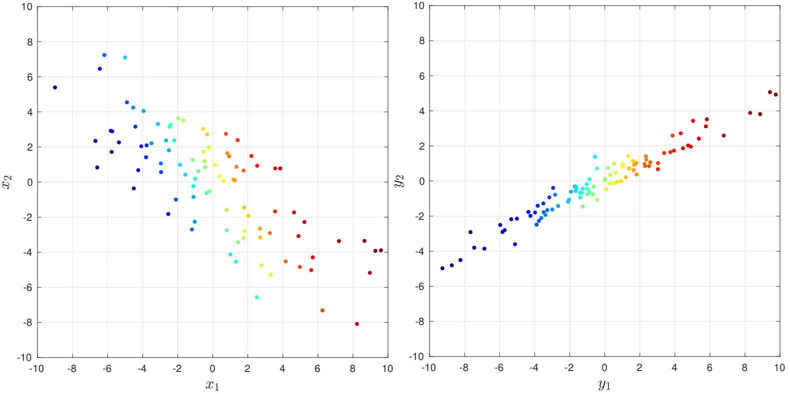
$$\rho_1 = 1.000 \tag{99}$$

$$\rho_2 = 0.211 \tag{100}$$

CCA Projection: Irrelevant, Orthogonal



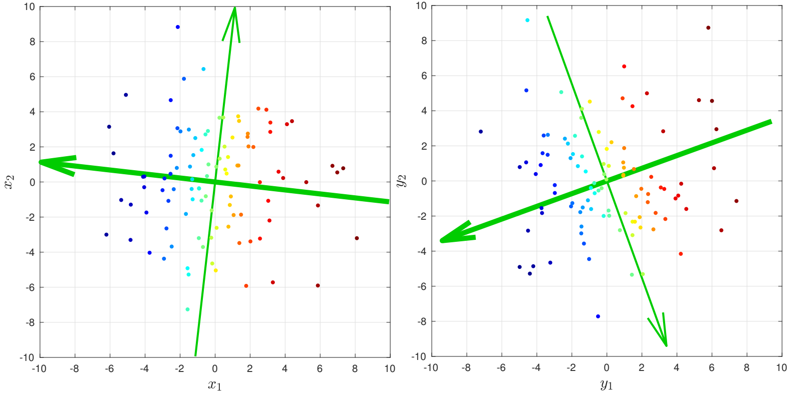
Point Set #2: Oblique Irrelevant Components



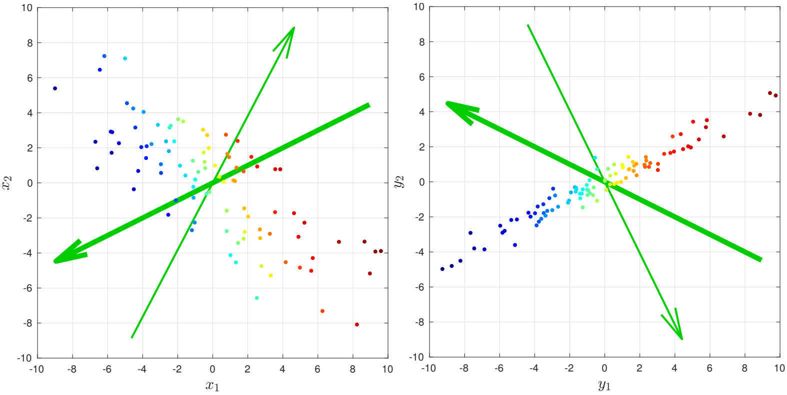
$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \begin{bmatrix} x_c \\ x_n \end{bmatrix} \qquad \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix} \begin{bmatrix} y_c \\ y_n \end{bmatrix}$$

$$x_n \sim \mathcal{N}(0, \sigma_x^2) \qquad y_n \sim \mathcal{N}(0, \sigma_y^2)$$

CCA in the Whitened Spaces: Oblique



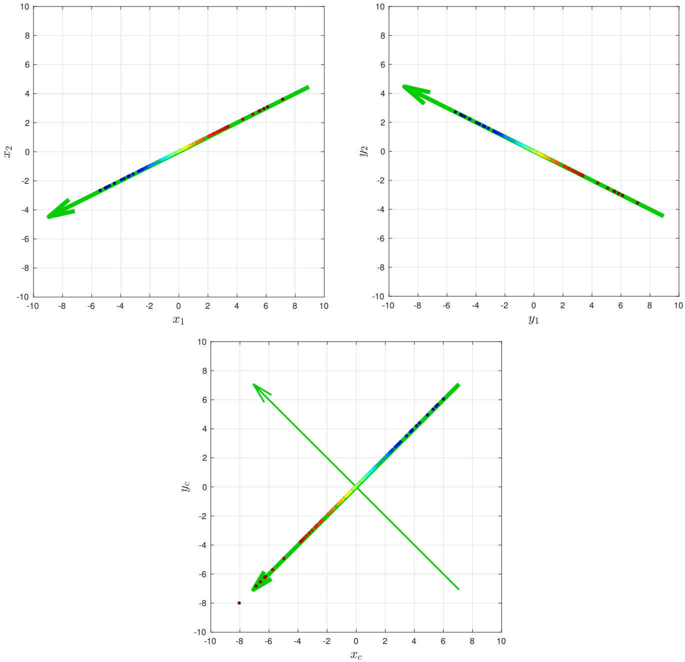
CCA in the Original Spaces: Oblique



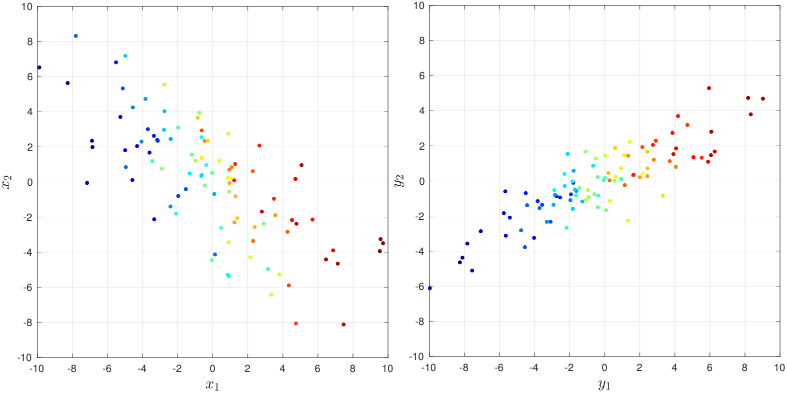
$$\rho_1 = 1.000 \tag{101}$$

$$\rho_2 = 0.211 \tag{102}$$

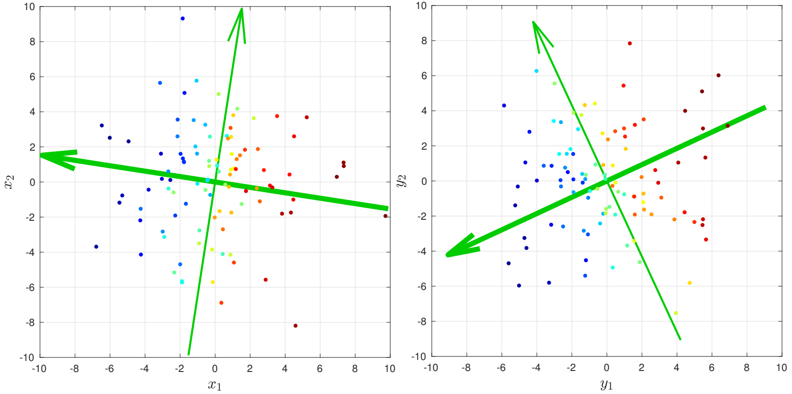
CCA Projection: Irrelevant, Oblique



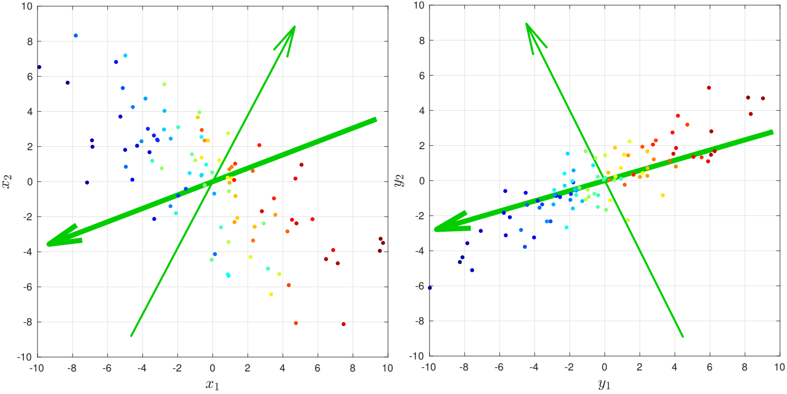
Point Set #3: Additive Noise in Each Space



CCA in the Whitened Spaces: Noise



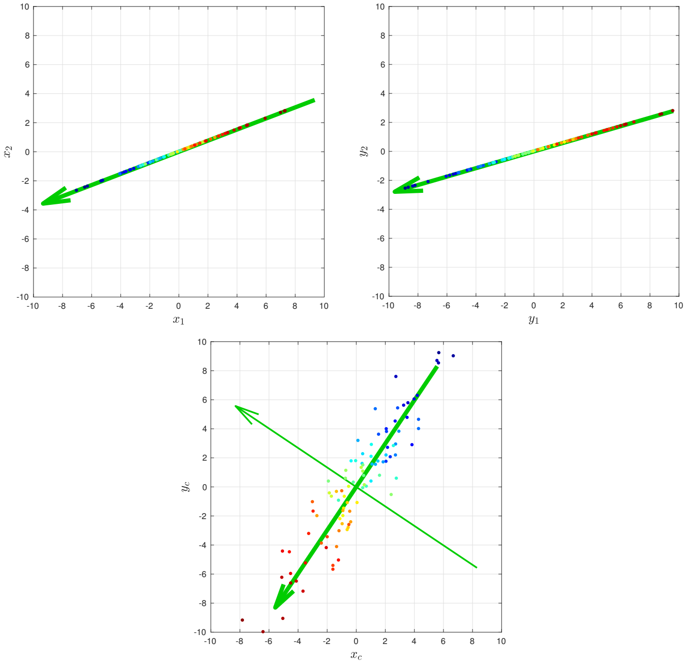
CCA in the Original Spaces: Noise



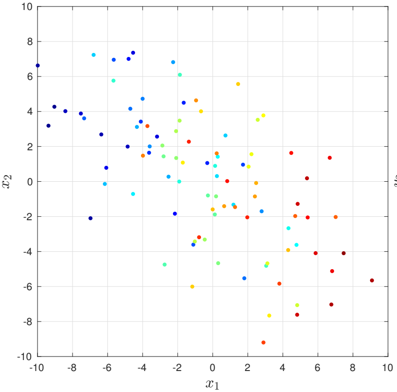
$$\rho_1 = 0.941 \tag{103}$$

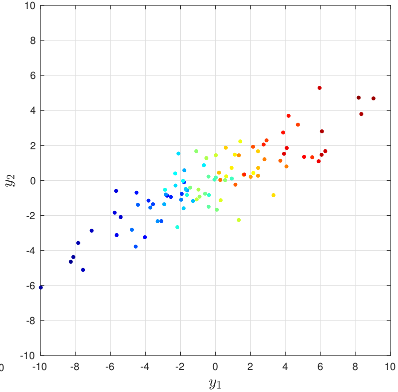
$$\rho_2 = 0.050 \tag{104}$$

CCA Projection: Additive Noise

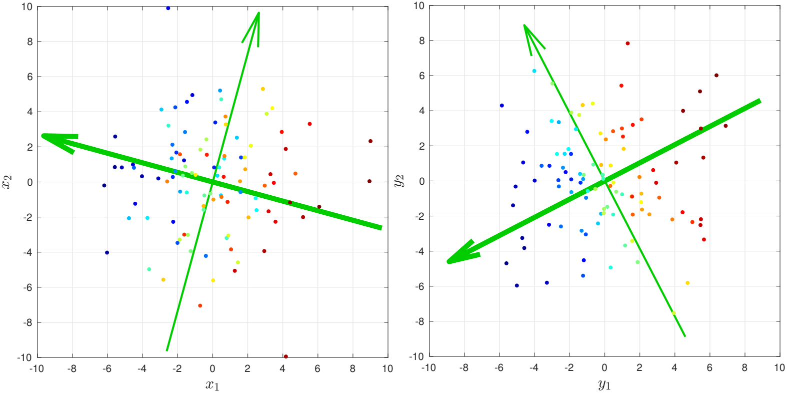


Point Set #4: Different Additive Noises

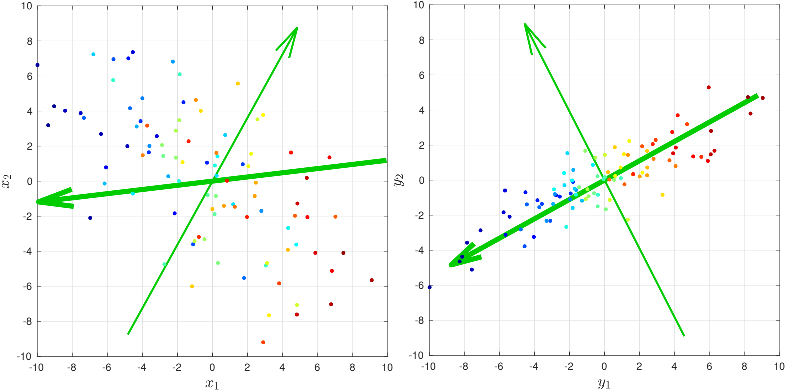




CCA in the Whitened Spaces: Different Noises



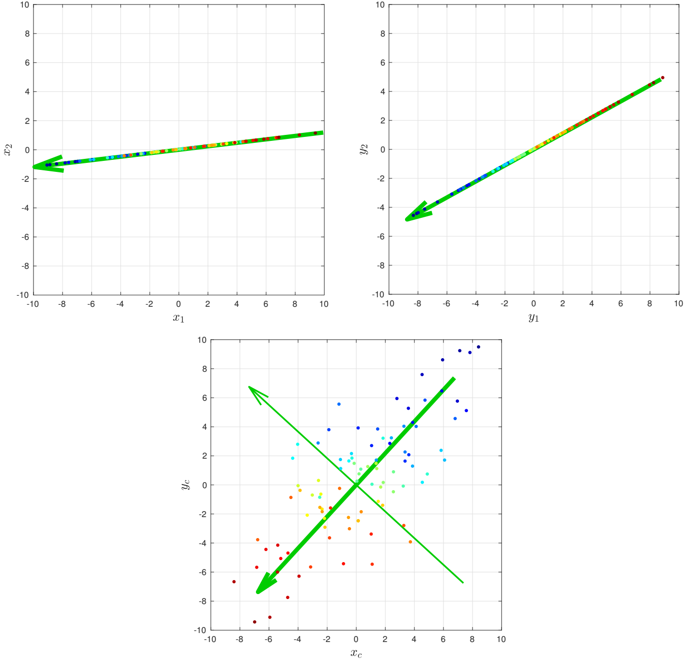
CCA in the Original Spaces: Different Noises



$$\rho_1 = 0.827 \tag{105}$$

$$\rho_2 = 0.077 \tag{106}$$

CCA Projection: Different Additive Noises



PCA vs CCA

- CCA studies the relationship between two sets, whereas PCA studies the relationship within a single set.
- CCA simultaneously find projection directions in the two spaces such that the projected data have maximal correlation, whereas PCA finds orthogonal projections that maximize variance in a single data space.
- CCA is limited to the minimal dimension of the two spaces.
- PCA and CCA are comptued using SVD of correlation matrices.
- CCA solution can be understood in three steps: whitening, alignment, and back projection.
- Pro: CCA is invariant with respect to scaling or general affine transformations of features.
- Con: A large correlation of CCA may not be significant. The significance can be tested statistically over random data pairing permutations.