Erasmus School of Economics

Weighted Linear Ridge Regression as an Approximation of Kernel Ridge Regression Kernels

MSc. Candidate: Li Huang

10 Dec. 2019

Erasmus University Rotterdam

Introduction

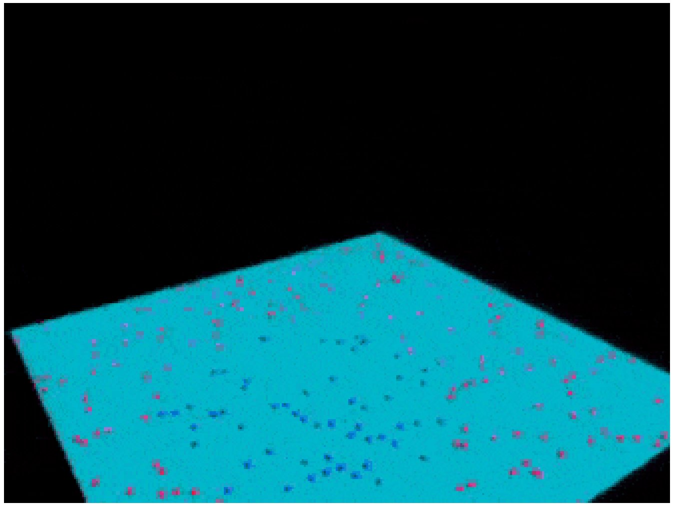
Model Types	Interpretable	Flexible	Model Names
Linear Model	☺	☹	{ OLS, ridge regression (RR) }
Nonlinear Model	☹	☺	{ kernel ridge regression (KRR) }
Linear Approximation	☺	☺	{ weighted ridge regression (WRR) }

Interpretable

$$\hat{y} = \alpha + \beta_1 x_1 + \beta_2 x_2$$

β_1 denotes the marginal effect of the predictor x_1

MC = MB



Source: udiprod (YouTube)

SVM

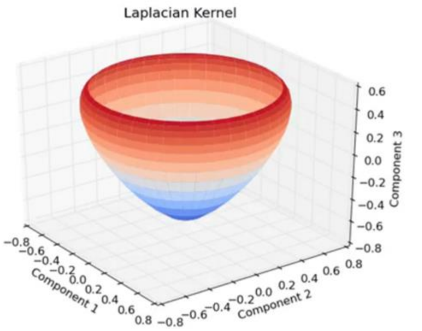
What is kernel function?

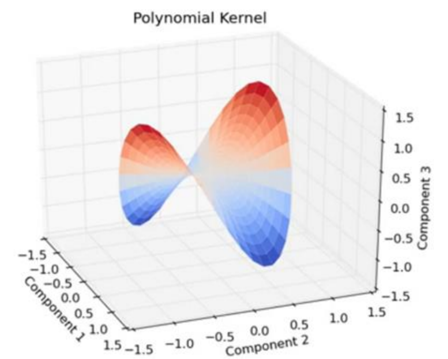
- Construct nonlinearity while saving computational complexity through calculating inner product.

$a = (a_1 + \ \dots + a_P)$

$\varphi(a) = (1, \sqrt{2}a_1, \dots, \sqrt{2}a_P, \ \ a_1^2, \dots, a_P^2, \ \sqrt{2}a_1 a_2, \dots, \sqrt{2}a_{P-1}a_P)'$

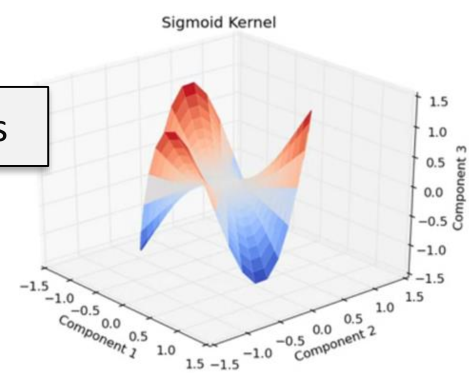
(infinity: RBF)





<u>Coordinates</u>: $\varphi(a)'\varphi(b)$ $\boxed{P^2 + 3P + \ 1 \text{ times}}$ *(infinity: RBF)*

<u>Inner product</u>: $\kappa(a, b) = \ \varphi(a)'\varphi(b) = (1 + a'b)^2$ $\boxed{2P + 1 \text{ times}}$



$K = \phi(X)\phi(X)'$

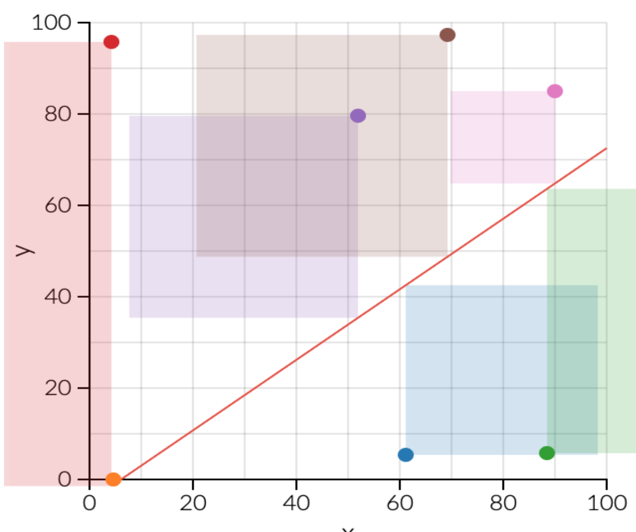
reduce computational complexity

Linear Models: OLS and Ridge Regression

1. **Ordinary Least Squares (OLS)**

$$min \parallel y - X\beta \parallel^2$$

$$\hat{y}_i = x_i' \boxed{(X'X)^{-1} y}$$



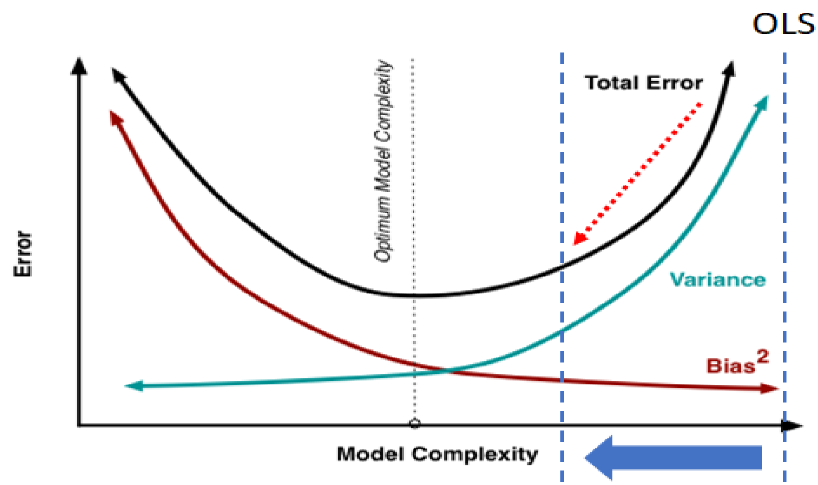
Source: DeepAI

2. **Ridge Regression (RR)**

$$MSE = \sigma^2 + u^2$$

$$min \parallel y - X\beta \parallel^2 + \lambda \parallel \beta \parallel^2$$

$$\hat{y}_i = x_i' \boxed{(X'X + \lambda I_P)^{-1} X'y}$$



Source: researchgate.net

Nonlinear Model: Kernel Ridge Regression

3. **Kernel Ridge Regression (KRR)**

$\emptyset(X) = Z$

Ridge regression: $\hat{y}_i = z_i'(Z'Z + \lambda I_M)^{-1} Z'y$

Z'Z is a (M x M) matrix

Duality of ridge regression: $\hat{y}_i = z_i' Z'(ZZ' + \lambda I_N)^{-1} y$

ZZ' is a (N x N) matrix

Kernel trick

$$\boxed{K = \emptyset(X)\emptyset(X)' = ZZ'}$$

$(N \times N)\ matrix$

$$k_i = Z z_i$$

$(N \times 1)\ vector$

$$\hat{y}_i = k_i'(K + \lambda I_N)^{-1} y$$

Weighted Ridge Regression

$Z = \emptyset(X) \approx S = XD_w$ (nonlinear $\rightarrow$ linear)

Euclidean distance: $\sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \cdots + (a_P - b_P)^2}$

- $\delta_{ij} = \sqrt{k_{ii} + k_{jj} - 2k_{ij}}$ $\quad Z$
- $d_{ij}(w) = \sqrt{\sum_{k=1}^{P} \left(x_{ik}w_{kk} - x_{jk}w_{kk}\right)^2}$ $\quad XD_w$

$\pi^2(w) = \sum_{i<j} \left(\delta_{ij} - d_{ij}(w)\right)^2$ SMACOF algorithm

(Scaling by MAjorizing a COmplicated Function)

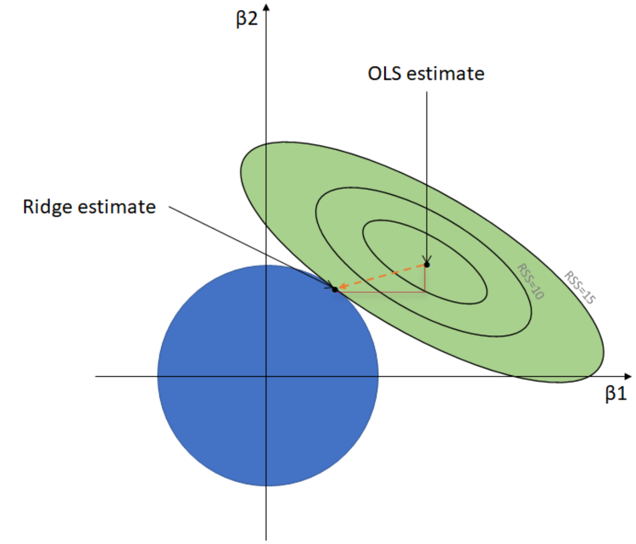
Iterative minimization procedure

Weighted Ridge Regression

RR:

$min \; \| y - X\beta \|^2 + \lambda \| \beta \|^2$

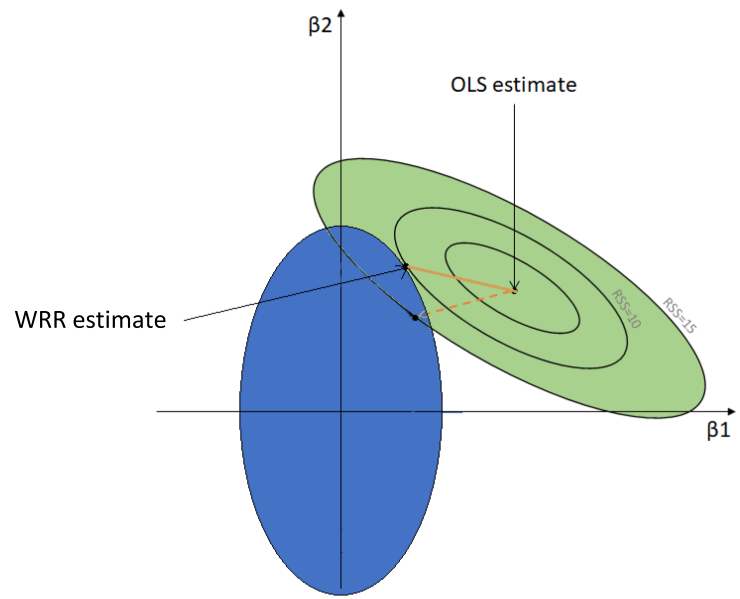
$\hat{y}_i = x_i'(X'X + \lambda I_P)^{-1}X'y$



WRR:

$min \; \| y - X\beta \|^2 + \lambda D_w^{-2} \| \beta \|^2$

$\hat{y}_i = x_i' \boxed{(X'X + \lambda D_w^{-2})^{-1}X'y}$



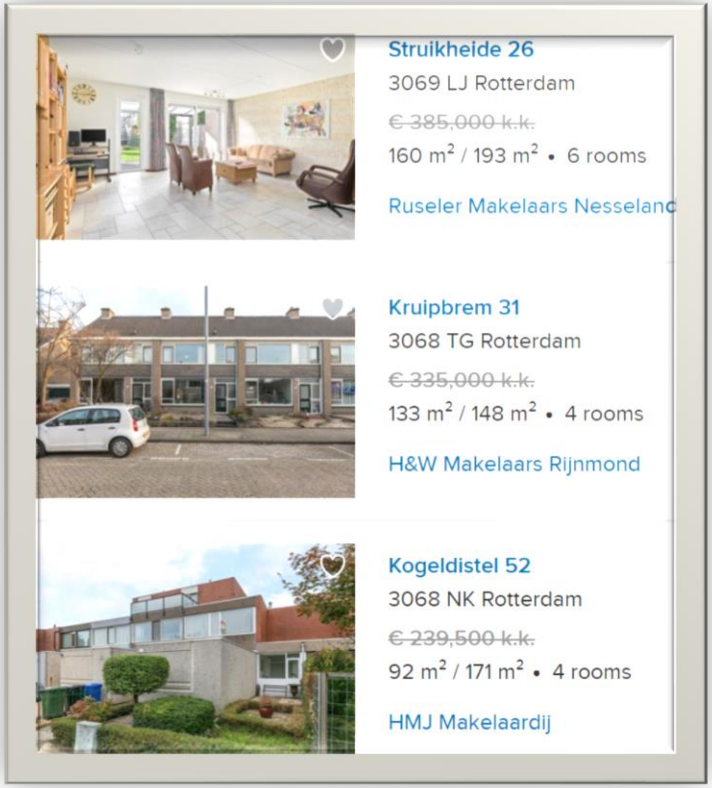
Case Study: Ommoord Housing Price

(Source: Google Maps)

Case Study: Ommoord Housing Price

Hedonic Price Function:
House characteristics & Environmental factors

House Characteristics	Enviromental Factors	Others
HouseAge	*Alexander100m*	*Term*
BckGardenSmt		*Index (HPI- NL)*
LivingArea		*IntRate*
CubicMeters		
nRooms	crime rate etc. are constant in this small area	
nResiLayers		
Full.ownership		
EnergyLabel		
nBedrooms		
PlotArea		



Source: funda.nl

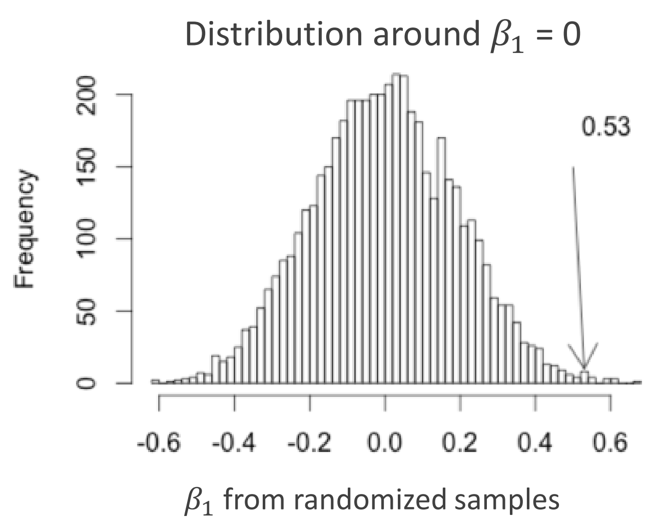
- ✓ 48 observations
- ✓ 13 out of 16 independent variables (AIC)

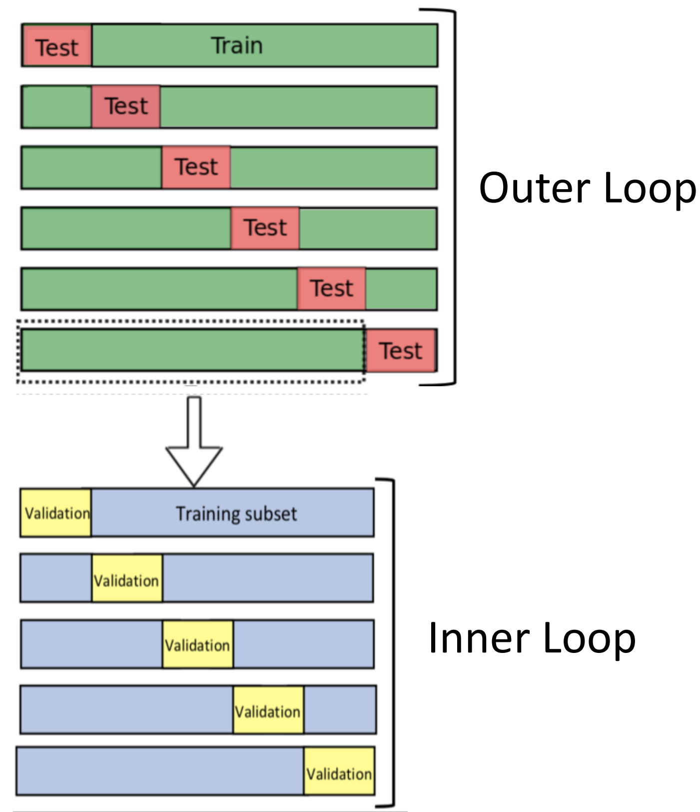
Predictive Power, NCV, and Significance Test

- Measurement of performance
  Root Mean Squared Error (RMSE)

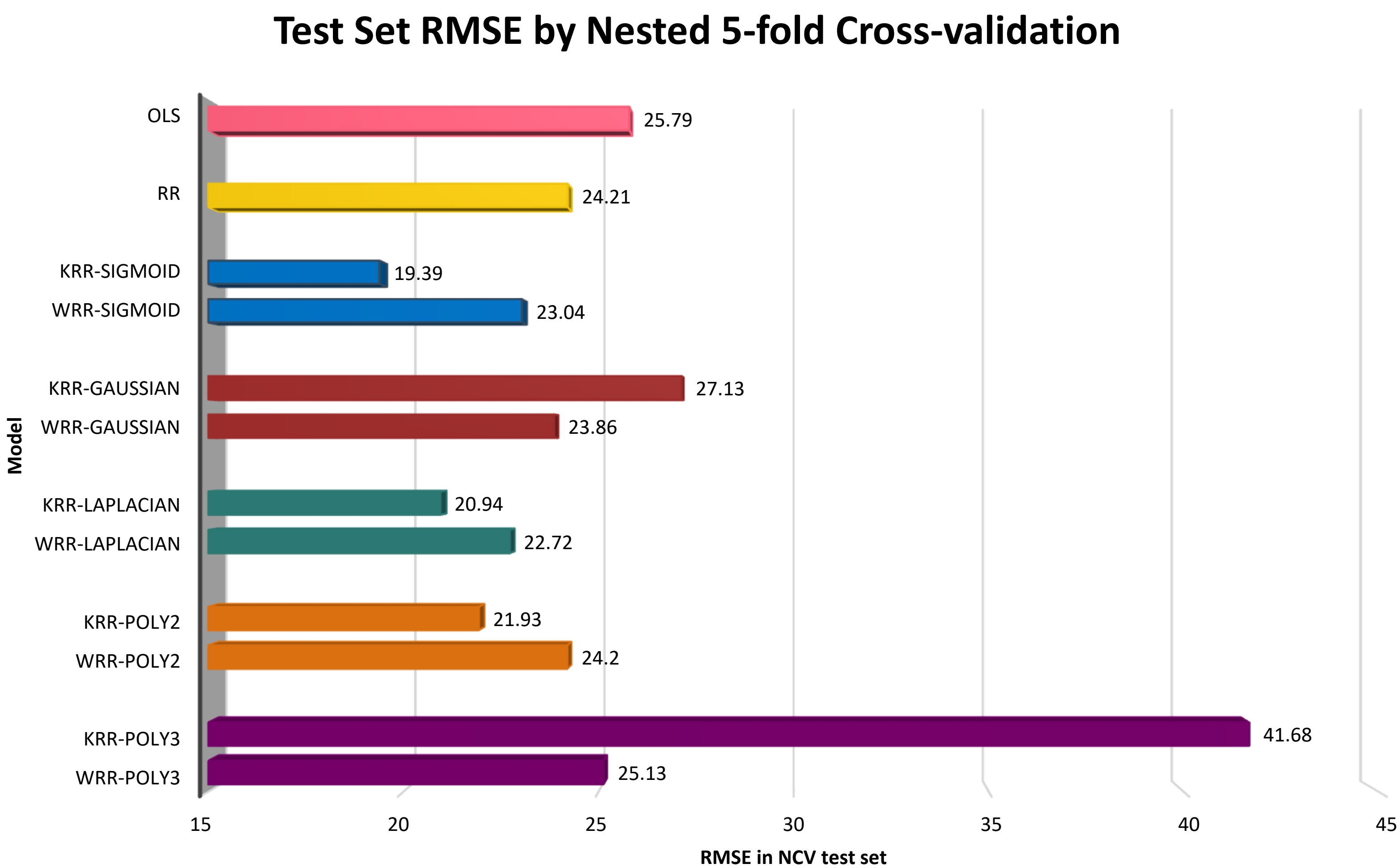
- 5-fold Nested Cross-validation
  - ❑ 6 test folds for outer loop
  - ❑ 5 validation fold in each inner loop

- Significance test
  - ❑ T-test for OLS
  - ❑ Permutation test for the rest models





Results - Predictive Performance



KRR-sigmoid (19.39) > WRR-sigmoid (23.04) > RR (24.21) > OLS (25.79);

KRR-Laplacian (20.94) > WRR-Laplacian (22.72) > RR > OLS;

KRR-polynom2 (21.93) > WRR-polynom2 (24.20) > RR > OLS.

Results - Interpretation of Predictors

	OLS regression					Ridge regression		WRR-sigmoid				WRR-Laplacian			
	coef.	p-value (t)		p-value (p)		**coef.**	p-value (p)	**coef.**	p-value (p)		D_w	**coef.**	p-value (p)		D_w
Index	**2.08**	0.274		0.109		**2.15**	0.912	**1.99**	0.783		0.224	**2.03**	0.798		0.271
Term	**1.03**	0.373		0.122		**1.35**	0.913	**0.92**	0.428		0.139	**1.08**	0.672		0.143
HouseAge	**-8.14**	0.030	**	0.082	*	**-7.53**	0.906	**-4.82**	0.910		0.109	**-5.18**	0.981		0.116
LivingArea	**12.25**	0.012	**	0.160		**9.70**	0.926	**9.99**	0.578		0.202	**8.64**	0.765		0.162
CubicMeters	**9.33**	0.031	**	0.409		**8.26**	0.909	**8.41**	0.020	**	-0.182	**8.99**	0.034	**	-0.184
nRooms	**-9.72**	0.014	**	0.278		**-7.36**	0.926	**-7.35**	0.959		0.165	**-7.40**	0.985		0.180
nResiLayers	**1.67**	0.331		0.140		**1.88**	0.908	**2.14**	0.506		-0.207	**2.44**	0.681		-0.197
BckGardenSmt	**2.90**	0.194		0.407		**3.22**	0.906	**3.14**	0.084	*	-0.173	**3.55**	0.172		-0.209
Alexander100m	**8.98**	0.030	**	0.054	*	**8.42**	0.915	**10.57**	0.004	***	0.242	**10.27**	0.005	***	0.226
Full.ownership	**6.35**	0.050	**	0.118		**4.62**	0.925	**5.15**	0.631		-0.240	**5.10**	0.832		-0.228
EnergyLabel_D	**-18.23**	0.034	**	0.244		**-9.45**	0.930	**-8.62**	0.443		0.217	**-10.01**	0.672		0.227
EnergyLabel_C	**0.18**	0.493		0.296		**6.95**	0.910	**8.09**	0.622		0.249	**6.29**	0.804		0.189
EnergyLabel_B	**-6.42**	0.164		0.145		**-1.22**	0.941	**-0.93**	0.382		0.124	**-1.39**	0.598		0.120
						$\lambda = 11.33$		$\lambda = 0.34$				$\lambda = 0.34$			

Conclusion

- $KRR > WRR > RR > OLS$

- Choice of kernels

- *Applications in Marketing Analytics:*
  features of a product → popularity
  features of a customer→ response rate in marketing campaign

- *Applications in Other Fields:*
  corporate finance (e.g. top manager's compensation)
  healthcare
  psychology
  …

Thank You