From learning with Fair regularizers to Physics aware models

Adrián Pérez-Suay

Department of Mathematics Education
Image Processing Laboratory
Universitat de València
Adrian.Perez@uv.es
http://isp.uv.es

Acknowledgements:

PID2019-109026RB-I00

Part I

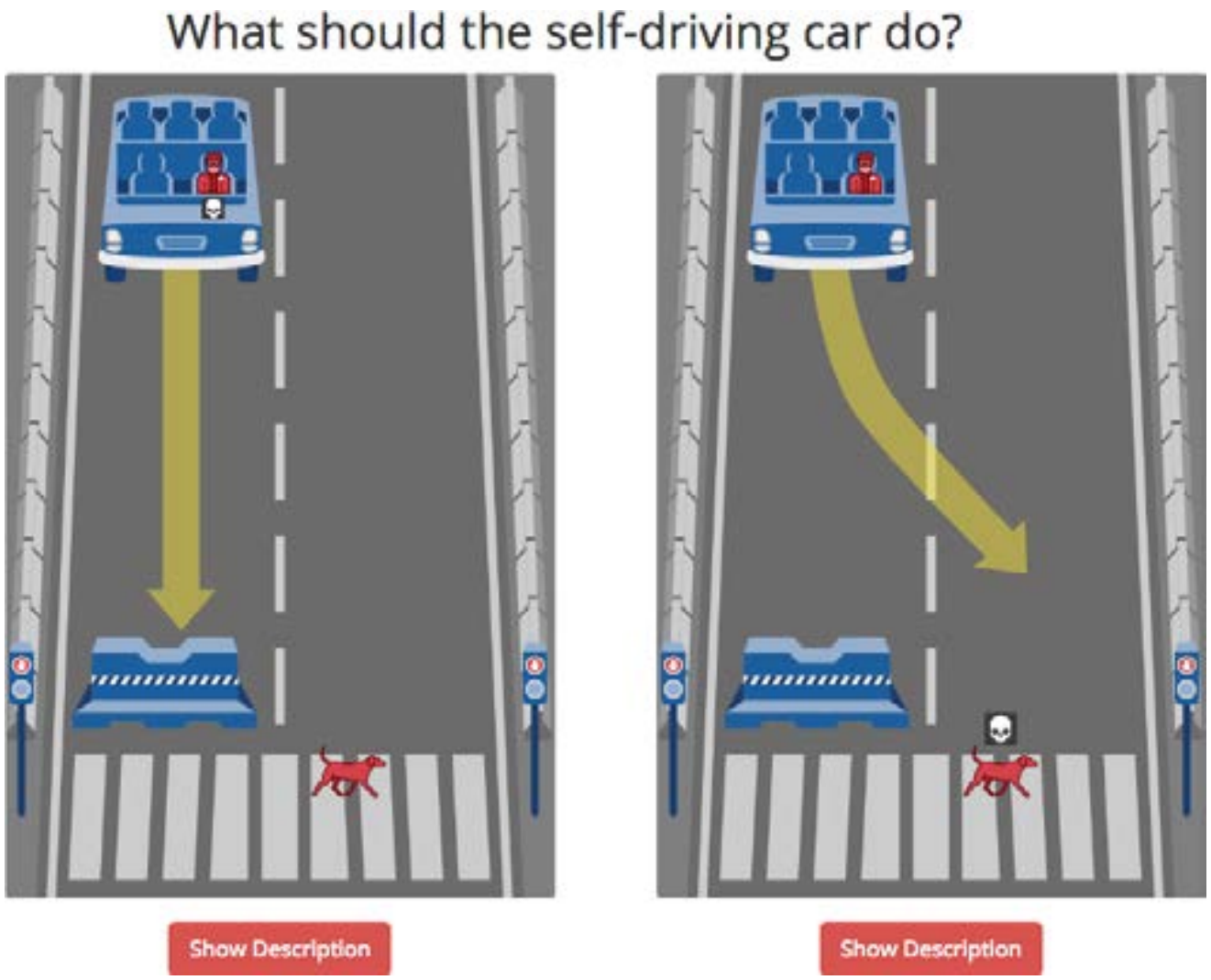
Fair Kernel Learning

Outline of part I

1. Motivation
2. Omitted-variable bias and fair learning
3. Fair learning regularization framework with kernels
   - Regression and classification
   - Dimensionality reduction
4. Experiments
   - Income prediction
   - Contraceptive adoption
5. Conclusions

1/ Motto: Self-driving car is almost solved!

1/ Motto: what if system is animalism bias?



1/ Motto: Learning from Tweeter

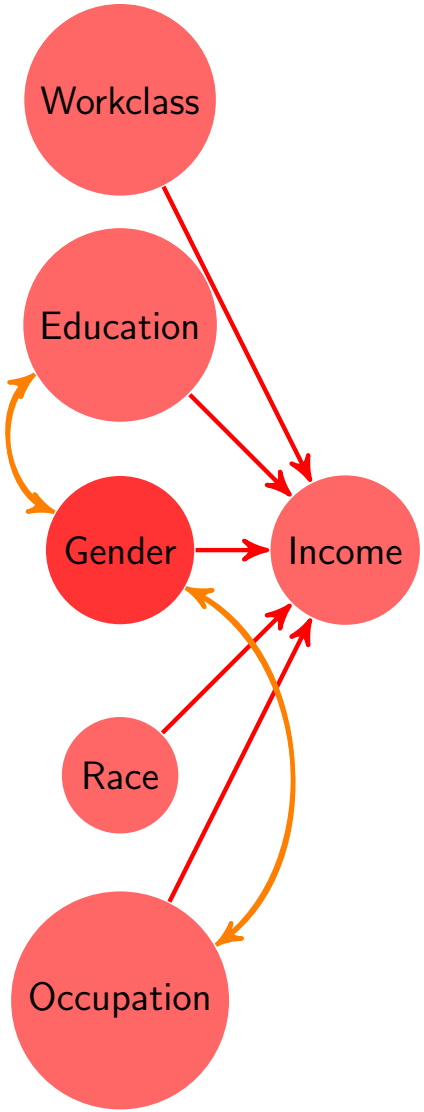
1/ Motto: Learning from Tweeter

1/ Motto: Learning from Tweeter

1/ Motto: Learning from Tweeter

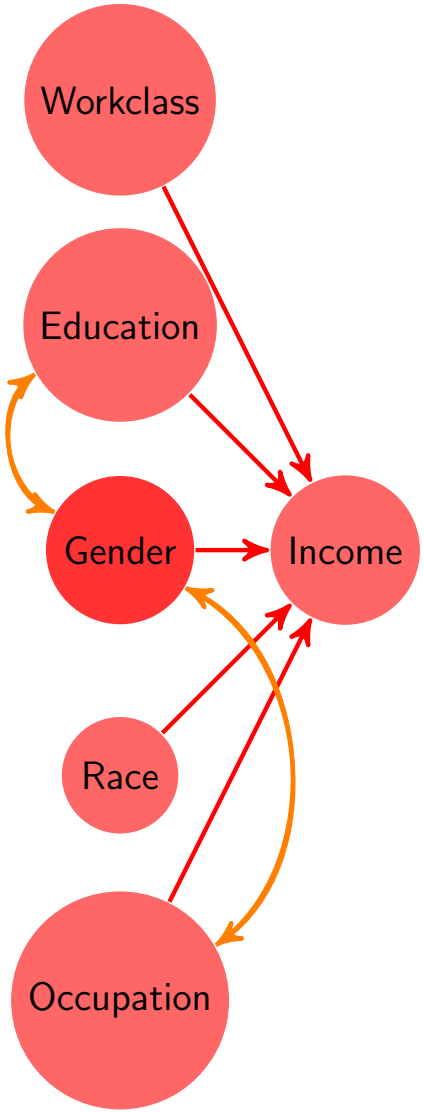
1/ Motto: Learning from Tweeter

2/ Fair learning (and the omitted variable bias)



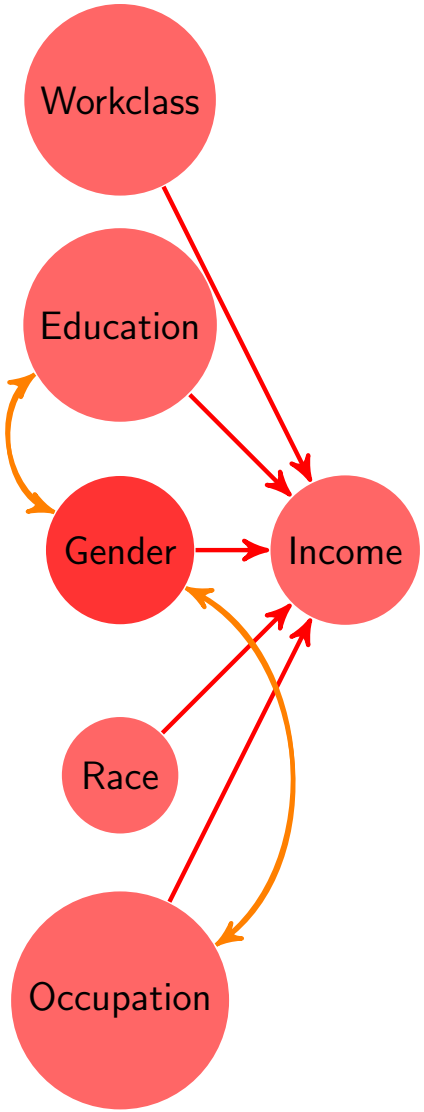
- Let's predict the income from some reasonable covariates
- Our company wants to be fair with the gender
- Removing the gender variable does not solve the problem
- Gender information is contained in other variables implicitly

2/ Fair learning (and the omitted variable bias)



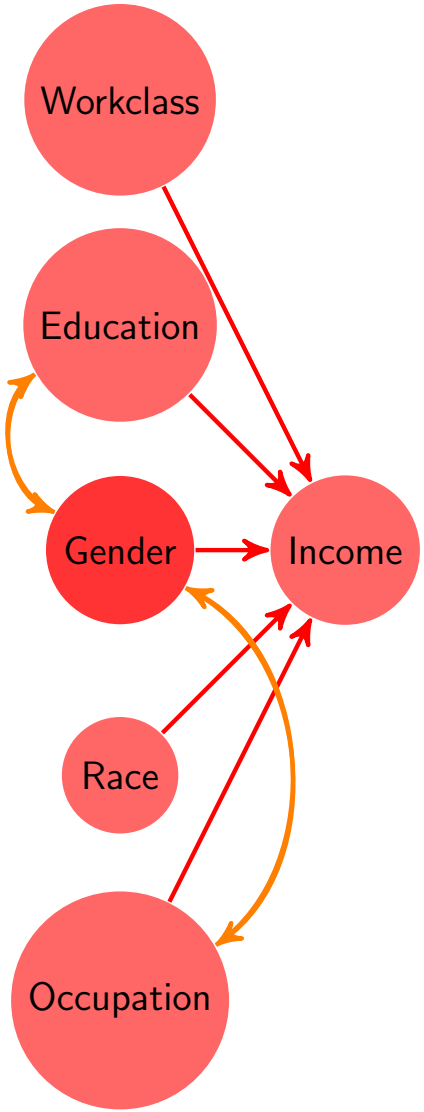
- Let's predict the income from some reasonable covariates
- Our company wants to be fair with the gender
- Removing the gender variable does not solve the problem
- Gender information is contained in other variables implicitly
- **How to avoid this omitted variable bias problem?**

2/ Fair learning setup



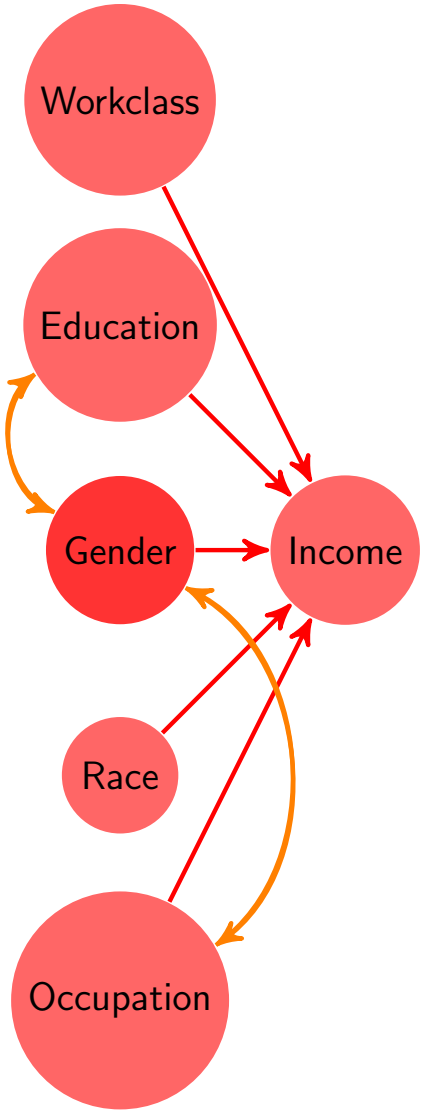
- **Goal:** Respect rules, ethics, laws; avoid disparate treatment

2/ Fair learning setup



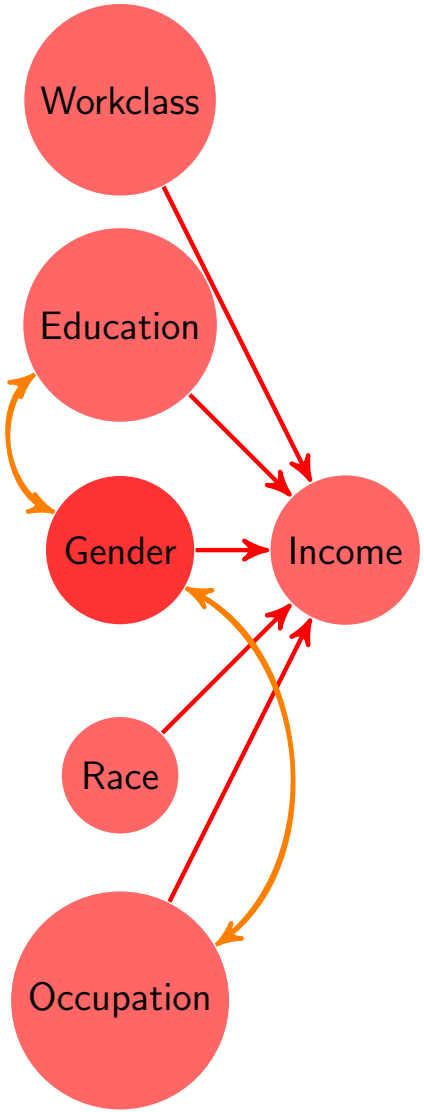
- **Goal:** Respect rules, ethics, laws; avoid disparate treatment
- **Premise:** Protected variables are worth including them

2/ Fair learning setup



- **Goal:** Respect rules, ethics, laws; avoid disparate treatment
- **Premise:** Protected variables are worth including them
- **Definition**:
  "A prediction is said to be totally fair with respect to the sensitive features $\mathbf{S}$ *if and only if* $\widehat{\mathbf{Y}} \perp \mathbf{S}$*."*

2/ Fair learning setup



- **Goal:** Respect rules, ethics, laws; avoid disparate treatment
- **Premise:** Protected variables are worth including them
- **Definition**:
  "A prediction is said to be totally fair with respect to the sensitive features $\mathbf{S}$ *if and only if* $\widehat{\mathbf{Y}} \perp \mathbf{S}$*."*
- **Idea:** Be accurate and insensitive to protected variables

3/ Regularization framework for fair prediction

$$\mathcal{L} = \frac{1}{n} \sum_{i=1}^{n} V(f(\mathbf{x}_i), \mathbf{y}_i) + \lambda \, \Omega(\|f\|_{\mathcal{H}}) + \mu \, I(f(\mathbf{x}), \mathbf{s})$$

3/ Regularization framework for fair prediction

$$\mathcal{L} = \frac{1}{n}\sum_{i=1}^{n} V(f(\mathbf{x}_i), \mathbf{y}_i) + \lambda\ \Omega(\|f\|_{\mathcal{H}}) + \mu\ I(f(\mathbf{x}), \mathbf{s})$$

- Linear model: $f := \hat{\mathbf{Y}} = \mathbf{XW}$
- L2 norm for the errors, $V := \|\mathbf{Y} - \hat{\mathbf{Y}}\|_2^2$
- Tikhonov's regularization for smoothness, $\Omega(\|f\|^2) := \|\mathbf{W}\|_2^2$
- Estimate independence I with mutual information [Kamishima, 2012]

3/ Regularization framework for fair prediction

$$\mathcal{L} = \frac{1}{n}\sum_{i=1}^{n} V(f(\mathbf{x}_i), \mathbf{y}_i) + \lambda\ \Omega(\|f\|_{\mathcal{H}}) + \mu\ I(f(\mathbf{x}), \mathbf{s})$$

- Linear model: $f := \hat{\mathbf{Y}} = \mathbf{XW}$
- L2 norm for the errors, $V := \|\mathbf{Y} - \hat{\mathbf{Y}}\|_2^2$
- Tikhonov's regularization for smoothness, $\Omega(\|f\|^2) := \|\mathbf{W}\|_2^2$
- Estimate independence I with mutual information [Kamishima, 2012]

Mutual Information

- ✗ Non Differentiable
- ✗ Depends on histograms estimation
- ✗ Unidimensional $d_s = d_f = 1$
- ✗ Difficult for continuous variables
- ✗ Computationally costly

3/ Regularization framework for fair prediction

$$\mathcal{L} = \frac{1}{n}\sum_{i=1}^{n} V(f(\mathbf{x}_i), \mathbf{y}_i) + \lambda\, \Omega(\|f\|_{\mathcal{H}}) + \mu\, \mathbf{HSIC}(\mathbf{f}(\mathbf{x}), \mathbf{s})$$

- Linear model: $f := \hat{\mathbf{Y}} = \mathbf{XW}$
- L2 norm for the errors, $V := \|\mathbf{Y} - \hat{\mathbf{Y}}\|_2^2$
- Tikhonov's regularization for smoothness, $\Omega(\|f\|^2) := \|\mathbf{W}\|_2^2$
- Estimate independence $\mathbf{I} = \mathbf{HSIC}$

Proposal: Hilbert-Schmidt Independence Criterion (HSIC)

- ✓ Differentiable (allows closed-form-solutions)
- ✓ Captures higher order relations
- ✓ Multidimensional: $d_s, d_f \geq 1$
- ✓ Allows: $d_s \neq d_f$
- ✓ Easy implementation!

3/ Regularization framework for fair prediction

$$\mathcal{L} = \|\mathbf{Y} - \hat{\mathbf{Y}}\|_2^2 + \lambda \ \|\mathbf{W}\|_2^2 + \mu \frac{1}{n^2} \mathrm{tr}(\hat{\mathbf{Y}}\hat{\mathbf{Y}}^\top \mathbf{S}\mathbf{S}^\top)$$

- Linear fair regression:

$$\widehat{\mathbf{W}} = (\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}} + \lambda \, \mathbf{I} + \frac{\mu}{n^2} \, \tilde{\mathbf{X}}^\top \tilde{\mathbf{S}} \tilde{\mathbf{S}}^\top \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^\top \mathbf{Y}, \quad \hat{\mathbf{Y}}_* = \mathbf{X}_* \widehat{\mathbf{W}}$$

3/ Regularization framework for fair prediction

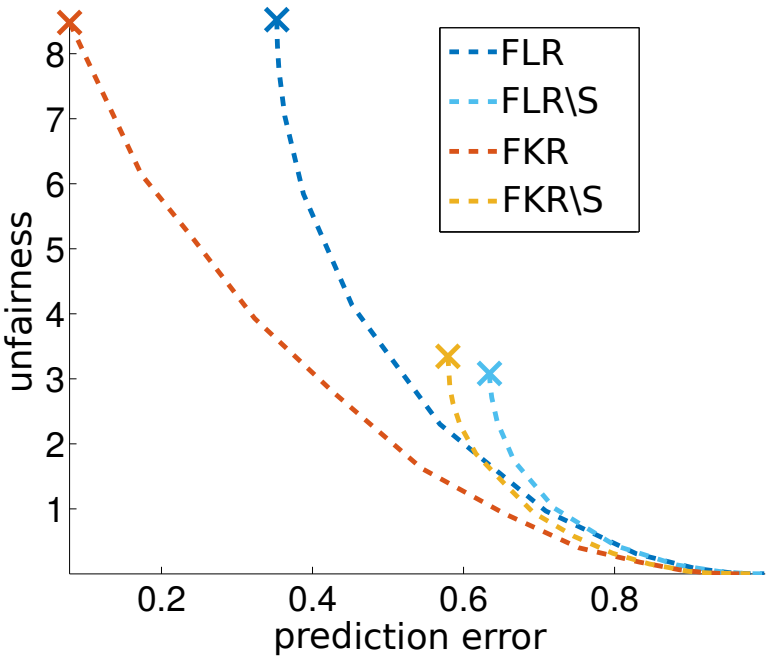
$$\mathcal{L} = \|\mathbf{Y} - \hat{\mathbf{Y}}\|_2^2 + \lambda \, \|\mathbf{\Phi}^\top \mathbf{\Lambda}\|_2^2 + \mu \tfrac{1}{n^2} \mathrm{tr}(\tilde{\mathbf{K}} \tilde{\mathbf{K}}_S)$$

- Linear fair regression:

$$\widehat{\mathbf{W}} = (\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}} + \lambda \, \mathbf{I} + \tfrac{\mu}{n^2} \, \tilde{\mathbf{X}}^\top \tilde{\mathbf{S}} \tilde{\mathbf{S}}^\top \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^\top \mathbf{Y}, \quad \hat{\mathbf{Y}}_* = \mathbf{X}_* \widehat{\mathbf{W}}$$

- Kernel (nonlinear) fair regression:

$$\widehat{\mathbf{\Lambda}} = (\tilde{\mathbf{K}} + \lambda \mathbf{I} + \tfrac{\mu}{n^2} \tilde{\mathbf{K}} \tilde{\mathbf{K}}_S)^{-1} \mathbf{Y}, \quad \hat{\mathbf{Y}}_* = \mathbf{K}(\mathbf{X}_*, \mathbf{X}) \widehat{\mathbf{\Lambda}}$$



- unfairness $= \mathrm{HSIC}(\hat{\mathbf{Y}}, \mathbf{S})$
- Removing the sensitive variable is worse
- Kernel better in accuracy than linear
- Fairness-accuracy nice tradeoffs

3/ Fair Dimensionality Reduction

Goal: Obtain fair feature representations

3/ Fair Dimensionality Reduction

Goal: Obtain fair feature representations

Notation:

- Input matrix $\mathbf{X} \in \mathbb{R}^{n \times d}$, Output matrix $\mathbf{Y} \in \mathbb{R}^{n \times c}$
- Sensitive: $\mathbf{S} \in \mathbb{R}^{n \times d_S}$

3/ Fair Dimensionality Reduction

Goal: Obtain fair feature representations

Notation:

- Input matrix $\mathbf{X} \in \mathbb{R}^{n \times d}$, Output matrix $\mathbf{Y} \in \mathbb{R}^{n \times c}$
- Sensitive: $\mathbf{S} \in \mathbb{R}^{n \times d_S}$

Linear: Fair projection matrix $\mathbf{U} \in \mathbb{R}^{d \times n_p}$, $n_p \leq d$

- $\mathbf{U}^* = \arg\max_{\mathbf{U}} \left\{ \frac{\text{HSIC}(\mathbf{XU}, \mathbf{X})}{\text{HSIC}(\mathbf{XU}, \mathbf{S})} \right\}$, $\mathbf{U} = [\mathbf{u}_1 | \cdots | \mathbf{u}_{n_p}]^\top$
- Solve gen.eig.: $\mathbf{C}_{xx}\mathbf{C}_{xx}^\top \mathbf{u} = \lambda \mathbf{C}_{xs}\mathbf{C}_{xs}^\top \mathbf{u}$

3/ Fair Dimensionality Reduction

Goal: Obtain fair feature representations

Notation:

- Input matrix $\mathbf{X} \in \mathbb{R}^{n \times d}$, Output matrix $\mathbf{Y} \in \mathbb{R}^{n \times c}$
- Sensitive: $\mathbf{S} \in \mathbb{R}^{n \times d_S}$

Linear: Fair projection matrix $\mathbf{U} \in \mathbb{R}^{d \times n_p}$, $n_p \leq d$

- $\mathbf{U}^* = \arg\max_{\mathbf{U}} \left\{ \frac{\text{HSIC}(\mathbf{XU}, \mathbf{X})}{\text{HSIC}(\mathbf{XU}, \mathbf{S})} \right\}$, $\mathbf{U} = [\mathbf{u}_1 | \cdots | \mathbf{u}_{n_p}]^\top$
- Solve gen.eig.: $\mathbf{C}_{xx}\mathbf{C}_{xx}^\top \mathbf{u} = \lambda \mathbf{C}_{xs}\mathbf{C}_{xs}^\top \mathbf{u}$

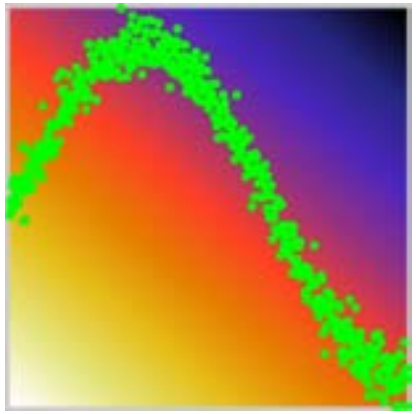
Kernel: Use representer theorem $\mathbf{U} = \mathbf{\Phi}^\top \mathbf{\Lambda}$

- $\mathbf{\Lambda}^* = \arg\max_{\mathbf{\Lambda}} \left\{ \frac{\text{Tr}(\mathbf{\Lambda}^\top \mathbf{\Phi}^\top \mathbf{\Phi}\mathbf{\Phi}^\top \mathbf{\Phi}\mathbf{\Lambda})}{\text{Tr}(\mathbf{\Lambda}^\top \mathbf{\Phi}^\top \mathbf{\Psi}\mathbf{\Psi}^\top \mathbf{\Phi}\mathbf{\Lambda})} \right\}$, $\mathbf{\Lambda} = [\boldsymbol{\alpha}_1 | \cdots | \boldsymbol{\alpha}_{n_p}]^\top$
- Solve gen.eig.: $\mathbf{K}_x \mathbf{K}_x \boldsymbol{\alpha} = \lambda \mathbf{K}_s \mathbf{K}_x \boldsymbol{\alpha}$

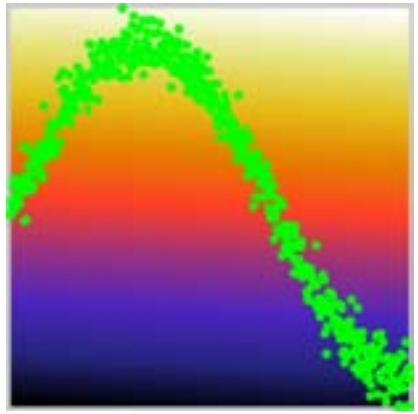
3/ Fair Dimensionality Reduction

Relation to invariance encoding!

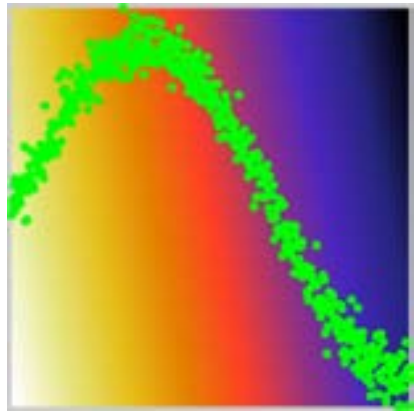
PCA



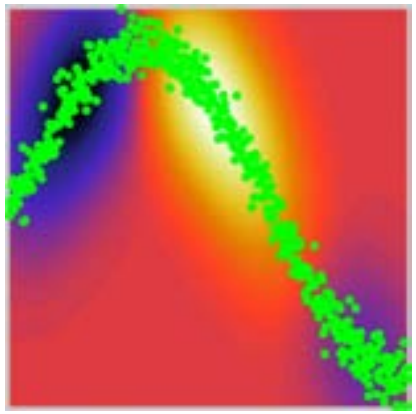
FDR ($\mathbf{S} = \mathbf{x}_1$)



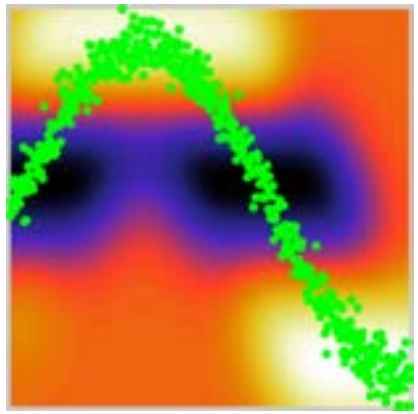
FDR ($\mathbf{S} = \mathbf{x}_2$)



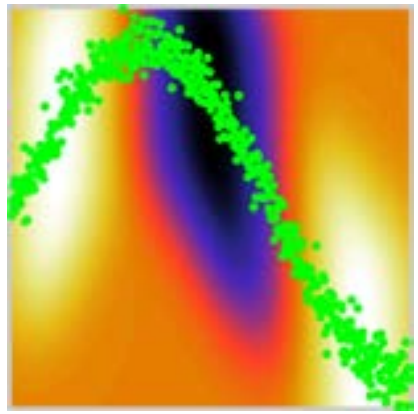
KPCA



KFDR ($\mathbf{S} = \mathbf{x}_1$)



KFDR ($\mathbf{S} = \mathbf{x}_2$)



4/ Experimental setup

- **UCI datasets**
  1. Adult (Classification, Dimensionality Reduction)
  2. Contraceptive (Classification)
- **Comparison of our proposals**
  - Standard versions (LR, PCA)
  - Linear and kernel versions
  - Naive solution of removing sensitive variables
- **Source code available**

  `http://isp.uv.es/soft_regression.html`

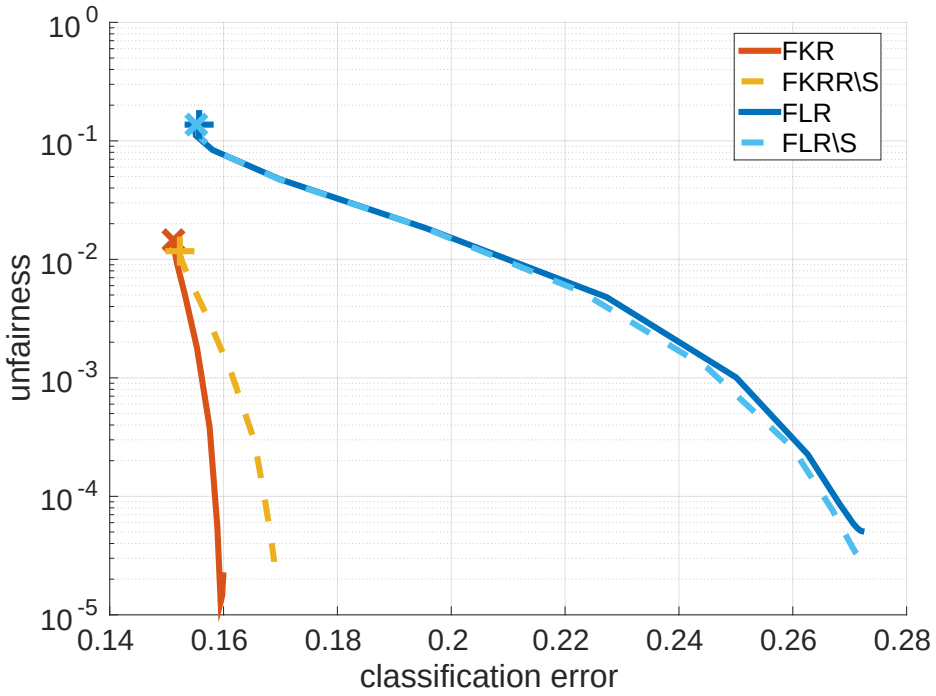
4/ Experiments: Adult dataset

- Income classification $\frac{50K}{year}\$$
- $n = 48842$
- UCI: $d = 14$ features
- `libsvm` (a9a): $d = 123$
- 25 trials

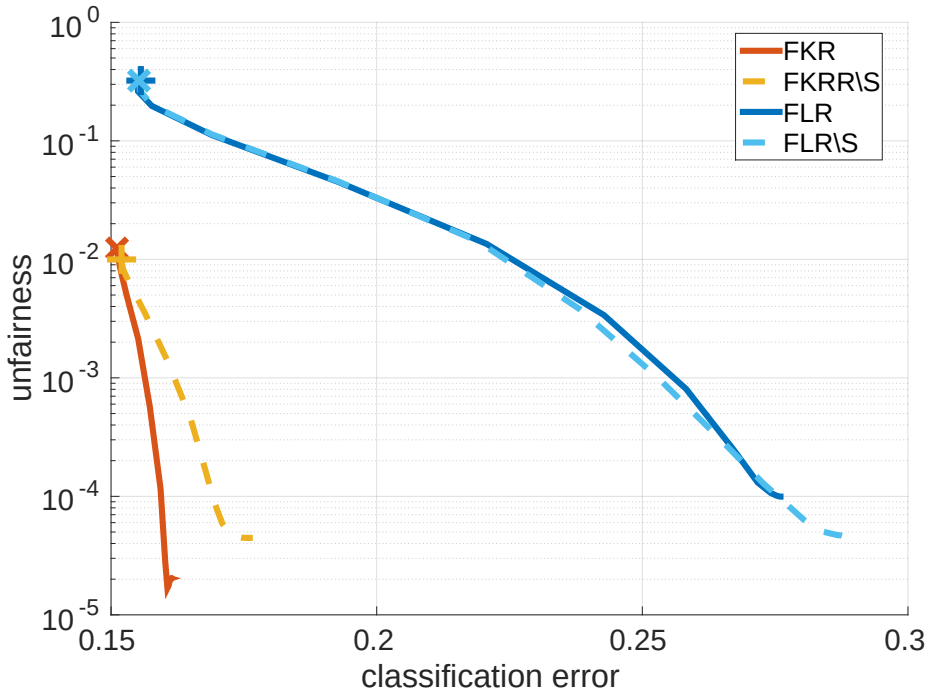
# feature	original feature
1	age
2	workclass
3	final weight
4	education
5	ed_num
6	marital_status
7	occupation
8	relationship
9	race
10	sex
11	capital_gain
12	capital_loss
13	hours × week
14	country

4/ Experiments: Adult dataset (class.)

$S = \{\text{sex}\}$



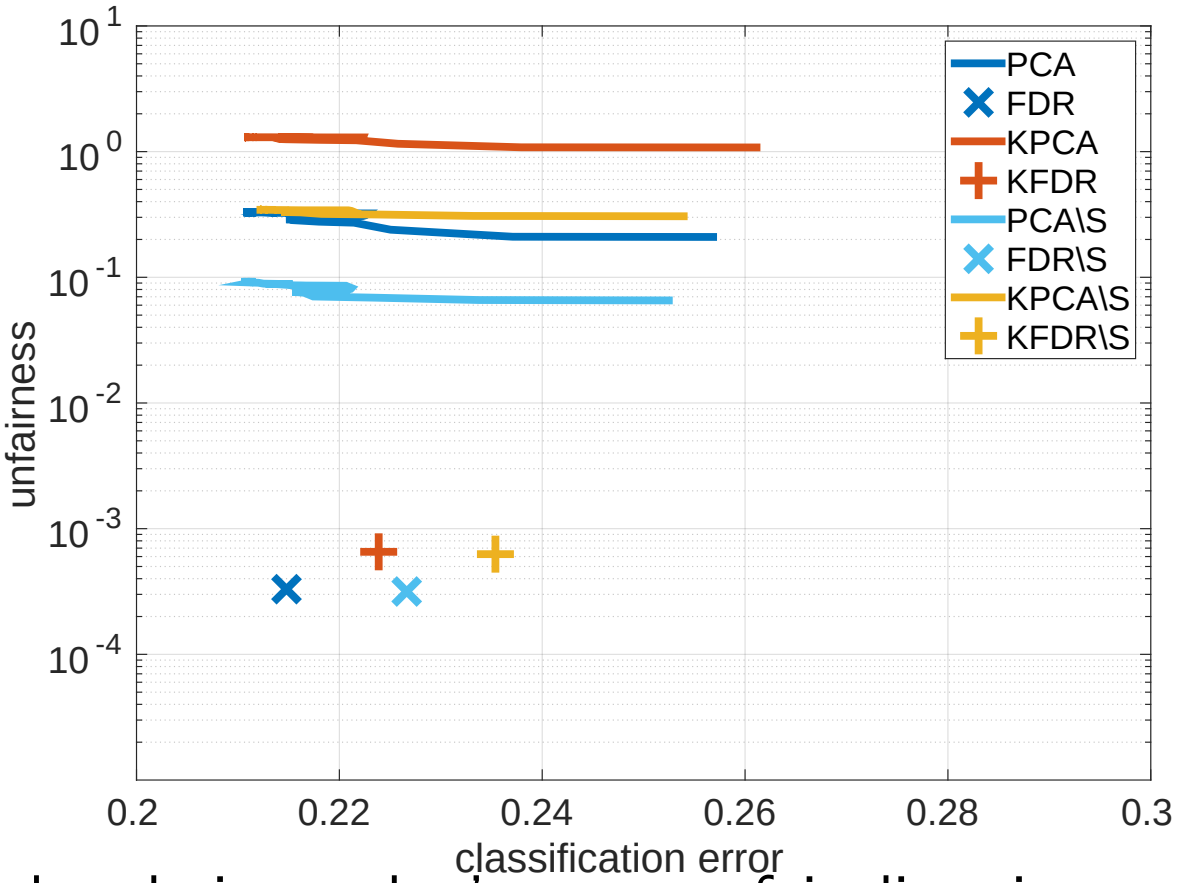
$S = \{\text{sex, race}\}$



- $(0, 0)$ means 0 error and fair
- Kernel less error and more fair than Linear
- It is better to use S

4/ Experiments: Adult dataset (dim.red.)

$S = \{\text{sex, race}\}$



- Standard techniques don't capture fair directions
- Fair D.R. obtain orders of magnitude more fair representations
- Fair D.R. dimensionality is limited to: $\text{rank}(\mathbf{K}\mathbf{K}_S)$

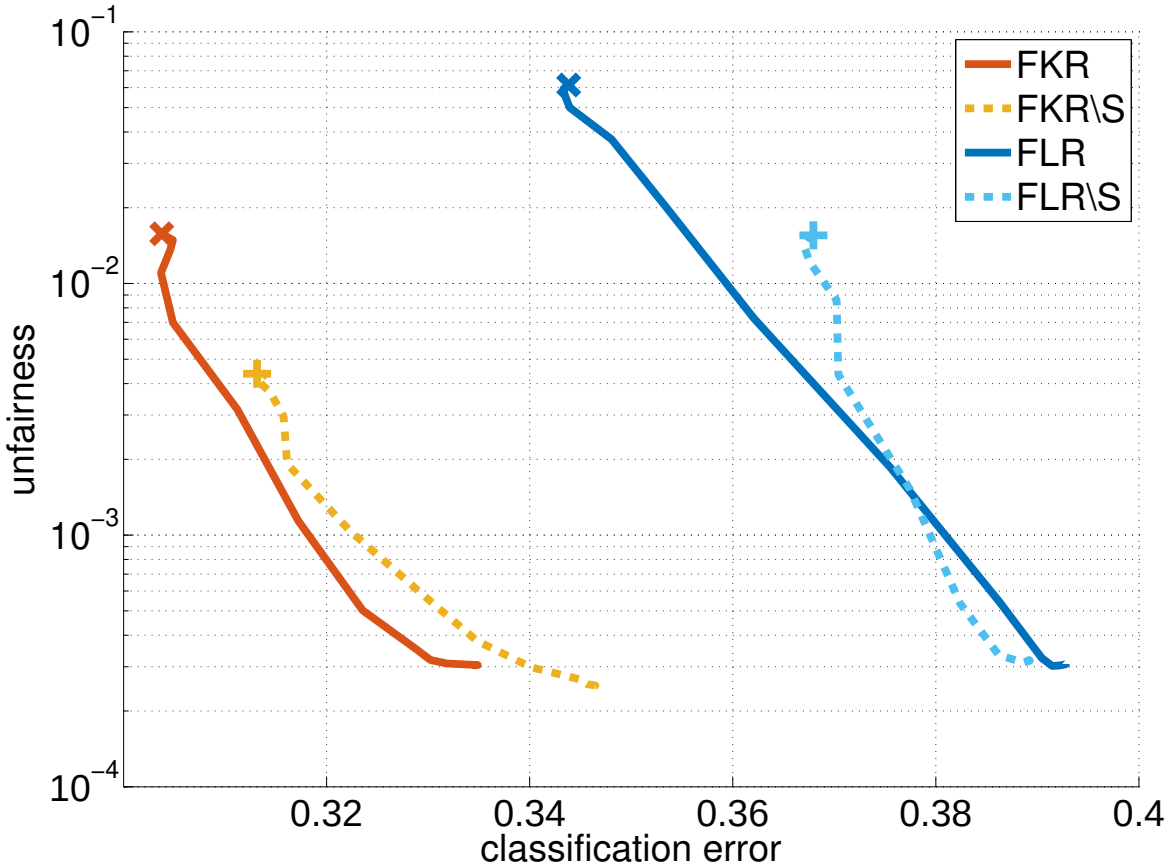
4/ Experiments: Contraceptive dataset

- Contraceptive method (yes/no)
- $n = 1473$
- UCI: $d = 9$ features
- 25 trials

# feature	feature description
1	wife's age
2	wife's education
3	husband's education
4	number of children ever born
5	wife's religion
6	wife's now working
7	husband's occupation
8	standard-of-living index
9	media exposure
10	contraceptive method used

4/ Experiments: Contraceptive dataset (class.)

$S = \{\text{wife's education}\}$



- $(0, 0)$ means 0 error and fair
- Kernel less error and more fair than Linear
- It is better to use S

5/ Conclusions (Part I)

Conclusions

- Fair kernel Prediction & Dim. Red.
- Closed form solutions & really easy implementation
- Consider as a General Fair Framework
- Better results (acc. & fairness) vs. removing sensitive S
- Tunable Trade-off: accuracy vs. fairness

5/ Conclusions (Part I)

Conclusions

- Fair kernel Prediction & Dim. Red.
- Closed form solutions & really easy implementation
- Consider as a General Fair Framework
- Better results (acc. & fairness) vs. removing sensitive S
- Tunable Trade-off: accuracy vs. fairness

Future work

- Comparison with state-of-the-art methods
- Real world data applications
- Deep Fair Learning

Part II

The Fair GP

Outline of part II

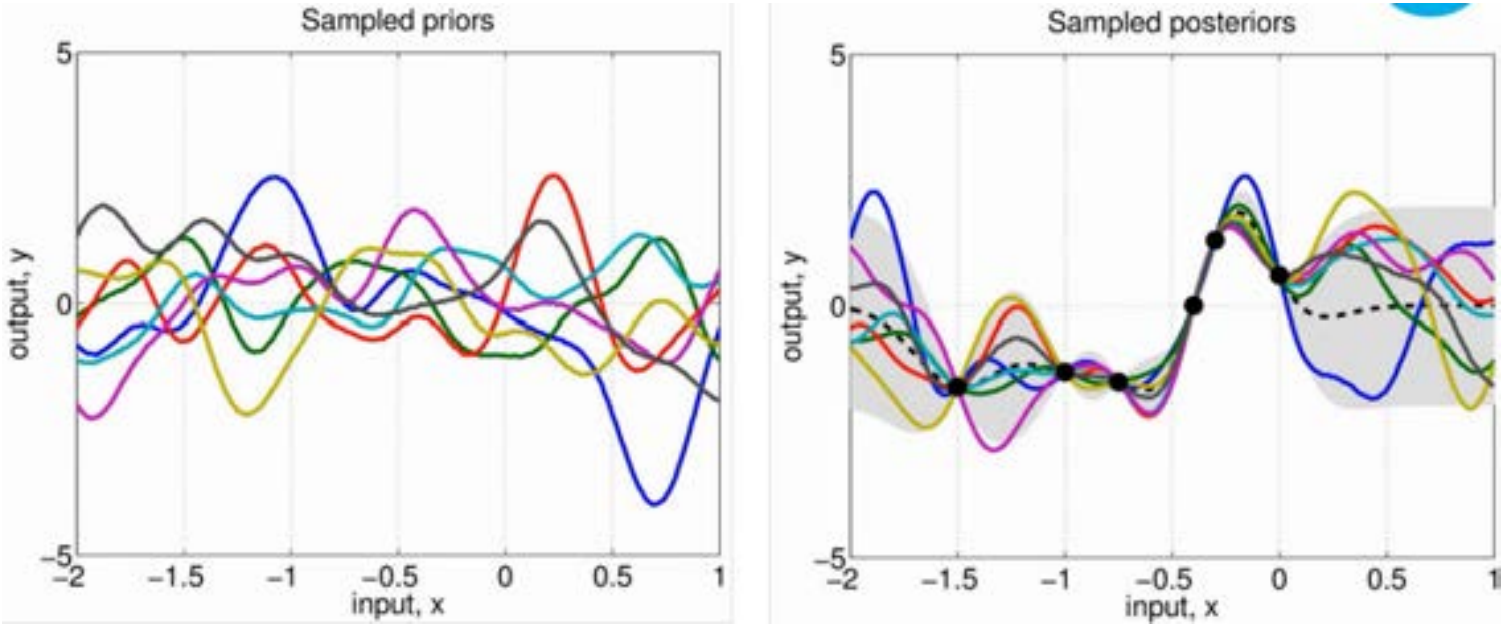
6. Motivation
7. GP models
8. Fair GP
9. Normalized dependence regularizers
10. Experiments
    - Toy dataset 1
    - Toy dataset 2
    - Crime and income prediction

1/ Motivation

Objectives

- Define a general framework of Empirical Risk Minimization with *fairness regularizers* and their interpretation
- Derive a Gaussian Process (GP) formulation of the fairness regularization framework
  - allows uncertainty estimation
  - hyperparameter selection
- Introduce a normalized version of the fairness regularizer
  - less sensitive to the choice of kernel parameters

2/ Gaussian Process models



- Observations y_i are assumed to arise from $p_\lambda(y_i|f(x_i))$
- GP regression assume normal likelihood
  $p_\lambda(y_i|f(x_i)) = \text{constant} - \frac{1}{2\lambda}(y_i - f(x_i))^2$
- Latent function $y_i = f(x_i) + \epsilon_i, \epsilon_i \sim \mathcal{N}(0, \lambda)$, iid
- Select model hyperparameters θ and λ by maximizing the marginal log-likelihood
- GP yield to a posterior distribution over unseen data $f(x_*)$

3/ Fair GP

Consider a particular instantiation of the above regularized functional

$$\min_{\beta \in \mathbb{R}^m} \left\{ -\sum_{i=1}^{n} \frac{\log p(y_i | f(x_i))}{\lambda} + \beta^\top \beta + \delta \beta^\top \Phi^\top \mathbf{HLH} \Phi \beta \right\}$$

Solution corresponds to the posterior mode in a Bayesian model using a **GP prior**:

$$f \sim \mathcal{GP}\left(0, k(\cdot,\cdot) - k_{\mathbf{X}\cdot}^\top (\mathbf{KHLH} + \delta^{-1}\mathbf{I})^{-1} \mathbf{HLH} k_{\mathbf{X}\cdot}\right).$$

where $k_{\mathbf{X}\cdot} = [k(\cdot, \mathbf{x}_1), \cdots, k(\cdot, \mathbf{x}_n)]^\top$, for any training set $\{\mathbf{x}_i\}_{i=1}^n$

4/ Fair Bayesian linear regressor

Empirical Risk Minimization

$$\begin{aligned}\boldsymbol{\beta}_* &:= \arg\min_{\boldsymbol{\beta}} \frac{1}{n}\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\|_2^2 + \frac{\lambda}{n}\|\boldsymbol{\beta}\|_2^2 + \eta\|\hat{\boldsymbol{\Sigma}}_{\mathrm{sx}}\boldsymbol{\beta}\|_2^2 \\ &= \arg\min_{\boldsymbol{\beta}} \frac{1}{\lambda}\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\|_2^2 + \boldsymbol{\beta}^\top(\mathbf{I}+\delta n^2\hat{\boldsymbol{\Sigma}}_{\mathrm{xs}}\hat{\boldsymbol{\Sigma}}_{\mathrm{sx}})\boldsymbol{\beta}.\end{aligned}$$

where $\hat{\boldsymbol{\Sigma}}_{\mathrm{sx}} = \frac{1}{n}\boldsymbol{\Psi}^\top H\boldsymbol{\Phi}$ is the empirical Cross-Covariance operator

Fair Bayesian linear regressor model

$$\begin{aligned}\mathbf{Y} &= \mathbf{X}\boldsymbol{\beta}+\epsilon, & \epsilon &\sim \mathcal{N}(0,\lambda), \\ \boldsymbol{\beta} &\sim \mathcal{N}(0,\boldsymbol{\Sigma}), & \boldsymbol{\Sigma} &= (\mathbf{I}+\delta n^2\hat{\boldsymbol{\Sigma}}_{\mathrm{xs}}\hat{\boldsymbol{\Sigma}}_{\mathrm{sx}})^{-1}\end{aligned}$$

5/ Fair Bayesian nonlinear regressor

Empirical Risk Minimization

$$\begin{aligned}\bar{\beta} &:= \text{argmin}\frac{1}{n}\|\mathbf{Y} - \mathbf{\Phi}_{\boldsymbol{x}}\boldsymbol{\beta}\|_2^2 + \frac{\lambda}{n}\|\boldsymbol{\beta}\|_2^2 + \eta\|\hat{\mathbf{\Sigma}}_{\text{sx}}\boldsymbol{\beta}\|_2^2 \\ &= \text{argmin}\frac{1}{\lambda}\|\mathbf{Y} - \mathbf{\Phi}_{\boldsymbol{x}}\beta\|_2^2 + \boldsymbol{\beta}^\top(\mathbf{I} + \delta n^2\hat{\mathbf{\Sigma}}_{\text{xs}}\hat{\mathbf{\Sigma}}_{\text{sx}})\boldsymbol{\beta}\end{aligned}$$

Fair Bayesian kernel regressor model

$$\begin{aligned} f \sim \mathcal{GP}(0, k^*(\cdot,\cdot)), \quad & y|f(\mathbf{x}) \sim \mathcal{N}(f(\mathbf{x}), \lambda), \\ k^*(\mathbf{x}, \mathbf{x}') = \langle\phi(\mathbf{x}), \mathbf{\Sigma}^*\phi(\mathbf{x}')\rangle, \quad & \mathbf{\Sigma}^* = (\mathbf{I} + \delta n^2\hat{\mathbf{\Sigma}}_{\text{xs}}\hat{\mathbf{\Sigma}}_{\text{sx}})^{-1}\end{aligned}$$

6/ Normalized dependence regularizers

By using the **normalized cross-covariance** operator $\mathbf{\Sigma}_{sx}$. Let $\hat{\mathbf{V}}_{sx} := \hat{\mathbf{\Sigma}}_{ss}^{-1/2}\hat{\mathbf{\Sigma}}_{sx}\hat{\mathbf{\Sigma}}_{xx}^{-1/2}$ Parameters of k is tuned from data, parameters from l are free to adjust

Empirical Risk Minimization

$$\bar{\beta} := \text{argmin}\left\{\frac{1}{n}\|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda\|\beta\|_2^2\right\} + \eta\|\hat{\mathbf{\Sigma}}_{ss}^{-1/2}\hat{\mathbf{\Sigma}}_{sx}\beta\|_2^2$$

Closed-form solution

$$\begin{aligned}\bar{\beta} &= (\mathbf{\Phi_x}^\top\mathbf{\Phi_x} + n\lambda\mathbf{I} + n\eta\hat{\mathbf{\Sigma}}_{xs}\hat{\mathbf{\Sigma}}_{ss}^{-1}\hat{\mathbf{\Sigma}}_{sx})^{-1}\mathbf{\Phi_x}^\top\mathbf{y} \\ &= \mathbf{\Phi_x}^\top(\mathbf{K} + n\lambda\mathbf{I} + \eta\mathbf{K}\tilde{\mathbf{L}}(\tilde{\mathbf{L}} + n\epsilon\mathbf{I})^{-1})^{-1}\mathbf{y}\end{aligned}$$

7/ Experiments: Toy dataset 1

- Sample x_1, x_2, z independently from $\mathcal{N}(0,1)$
- assuming z unobserved, let the sensitive variable be $x_3 = \frac{1}{\sqrt{2}}(x_1 + z)$
- x_1 and x_3 are correlated. Let the true function of interest be

$$f(\mathbf{x}, z) = \text{sign}((x_1 - z)x_3)|x_2|,$$

where $\mathbf{x} = [x_1, x_2, x_3]^T$

- We now further assume that **the observations y include a bias that is based on the sensitive variable x_3**

$$y = f(\mathbf{x}, z) + 2b\mathbf{1}_{\{x_3>0\}} - b + \epsilon,$$

i.e. the observations are on average increased by b when $x_3 > 0$ and decreased by b otherwise.

$$\tilde{x}_1 = x_1 - \frac{\mathbb{E}[x_1 x_3]}{\mathbb{E}x_3^2} x_3 = x_1 - \frac{1}{\sqrt{2}} x_3 = \frac{1}{2}(x_1 - z). \qquad (1)$$

7/ Experiments: Toy dataset 1

Table: The R^2 wrt. observations (left) and wrt. true value (right).

Approach	KRR	GPR	KRR	GPR
Standard	0.606 ± 0.002	0.612 ± 0.002	0.332 ± 0.003	0.356 ± 0.003
$\eta = 2 \times 10^{-3}$	0.600 ± 0.002	0.610 ± 0.001	0.358 ± 0.003	0.335 ± 0.002
$\eta = 0.2$	0.567 ± 0.001	0.586 ± 0.009	0.341 ± 0.005	0.394 ± 0.010
$\eta = 20$	0.488 ± 0.008	0.506 ± 0.012	0.466 ± 0.011	0.472 ± 0.008
$\eta = 200$	0.384 ± 0.011	0.403 ± 0.005	0.321 ± 0.014	0.530 ± 0.004
OSV	0.238 ± 0.007	0.196 ± 0.013	0.123 ± 0.008	0.098 ± 0.019
FRL	-0.021 ± 0.002	-0.009 ± 0.001	-0.024 ± 0.002	-0.011 ± 0.001

Table: The correlation between $\hat{y}$ and x_3.

Approach	KRR	GPR
Standard	0.3917 ± 0.0011	0.3863 ± 0.0013
$\eta = 2 \times 10^{-3}$	0.4053 ± 0.0019	0.3853 ± 0.0024
$\eta = 0.2$	0.3337 ± 0.0104	0.3257 ± 0.0206
$\eta = 20$	0.1364 ± 0.0150	0.2234 ± 0.0455
$\eta = 200$	0.1066 ± 0.0078	0.0139 ± 0.0031
OSV	0.2976 ± 0.0053	0.3195 ± 0.0058
FRL	-0.0010 ± 0.0012	-0.0102 ± 0.0013

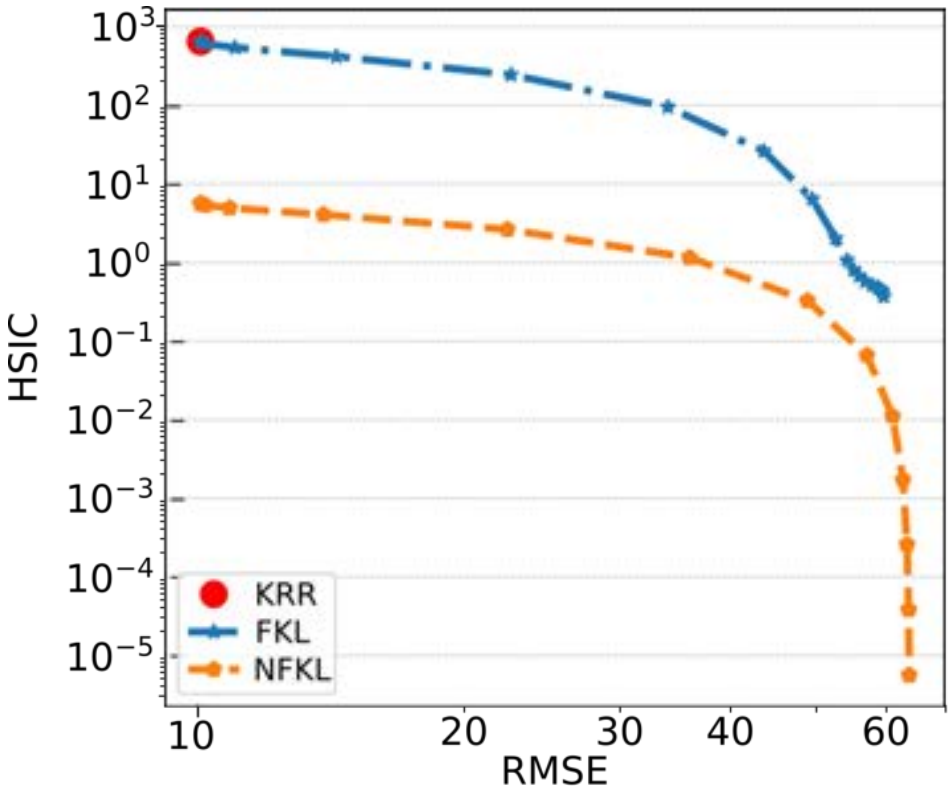
$R^2 = \frac{\text{variance explained by predictor}}{\text{total variance}}$ OSV: Omission Sensitive Variable

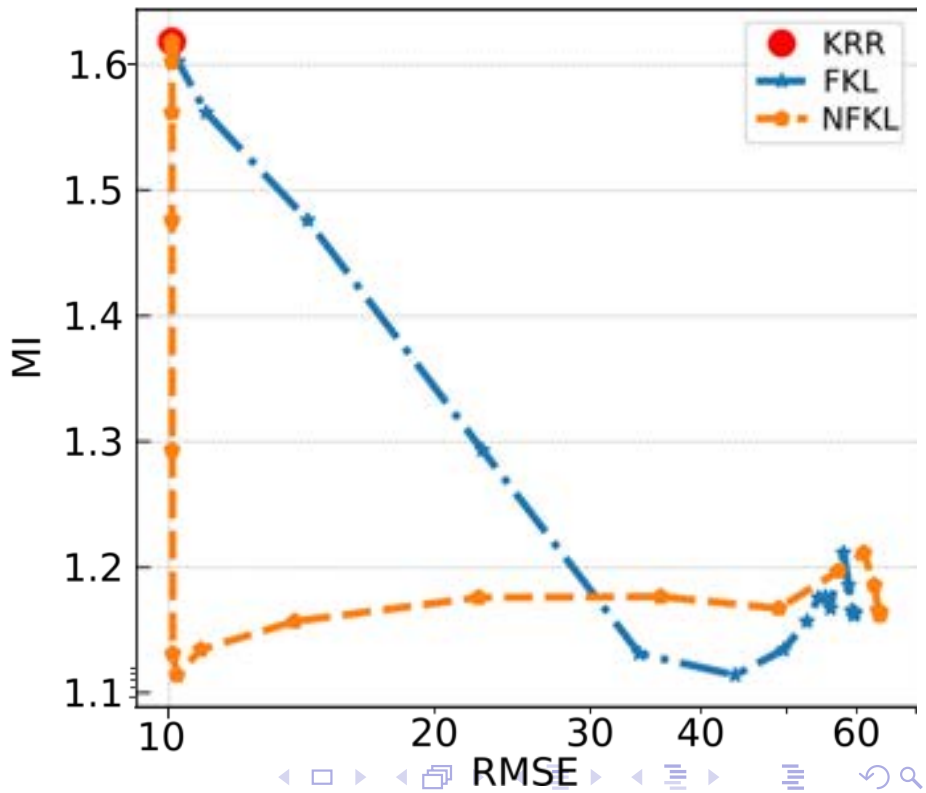
FRL: Fair Representation Learning (Learning fair representations, Zemel, ICML2013)

7/ Experiments: Toy dataset 2

We next consider a simple simulated dataset following the model from Fair Kernel Learning paper

$$y = x^2 + s^2 + \epsilon, \qquad x|s \sim \mathcal{N}(\log(|s|), \sigma_x^2),$$
$$s \sim \mathcal{N}(0, \sigma_s^2), \qquad \epsilon \sim \mathcal{N}(0, \sigma_y^2).$$

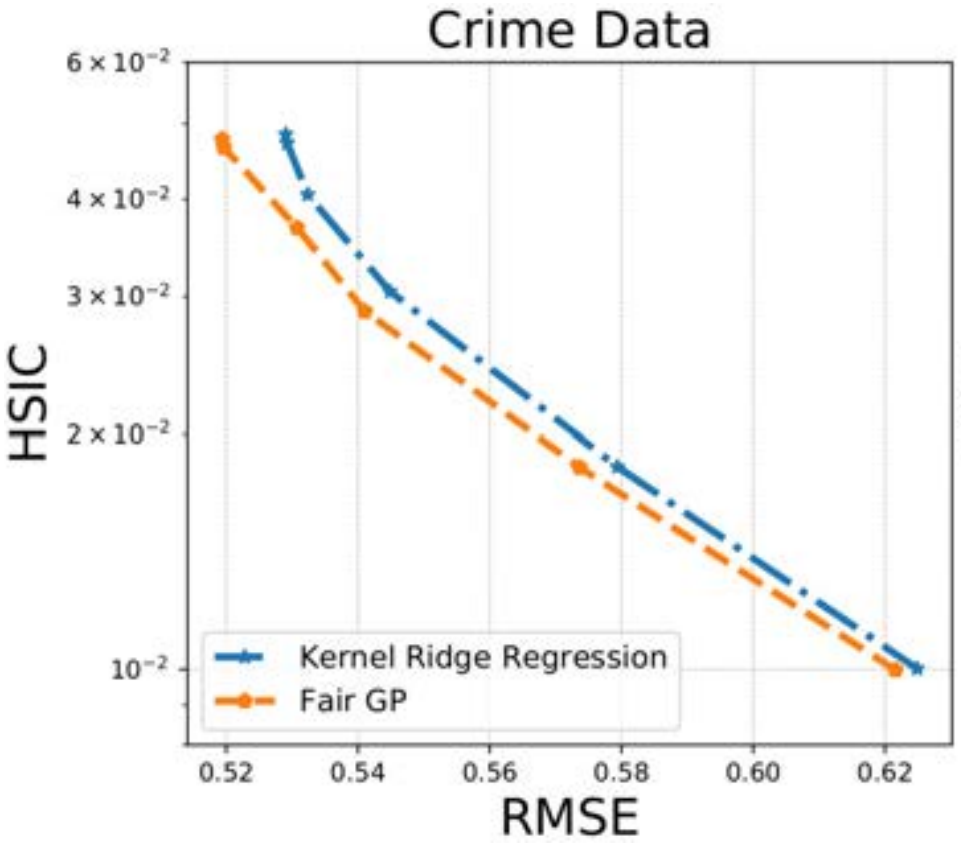


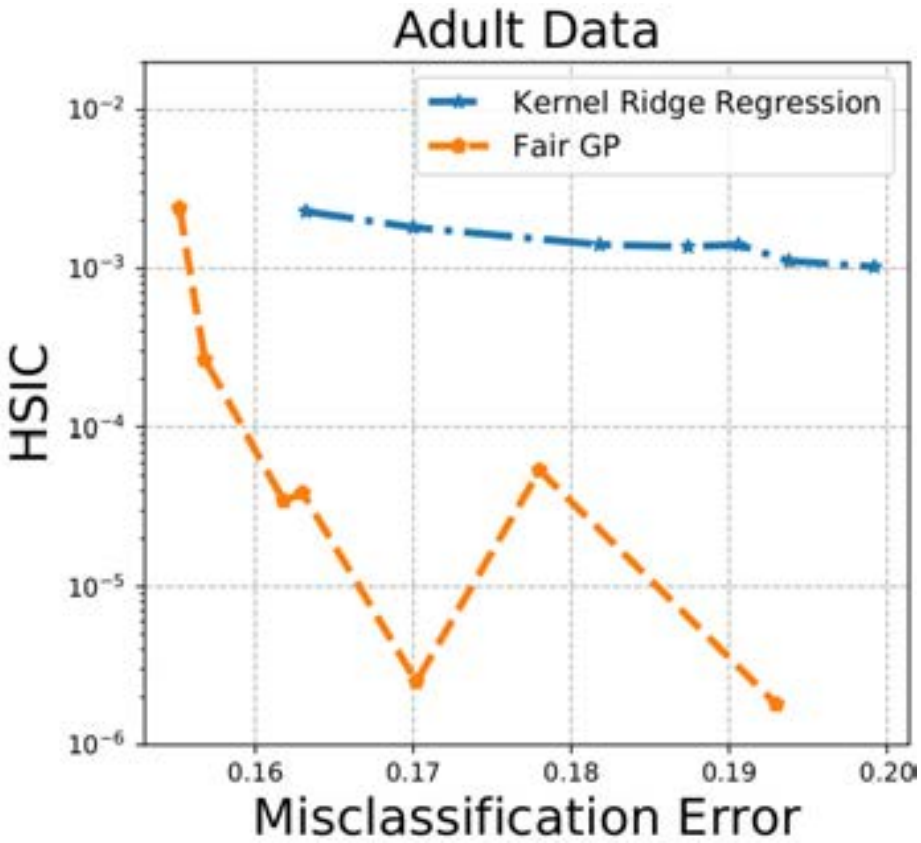


7/ Experiments: Crime and income prediction

Communities and Crime $\rightarrow$ predict crime rate USA, race is sensitive

Adult Income dataset





7/ Experiments: Crime and income prediction

Fair GP with ARD Kernel:

for $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^d$, $k(\mathbf{x}, \mathbf{x}') = \exp\left(-\sum_{i=1}^{d} \theta_i^{-2}(x_i - x_i')^2 \right)$

Table: The change of θ_i for sensitive variables with and without fair lerning

Sensitive Variable	GP	Fair GP
Race-Black	1.809 ± 0.216	2.939 ± 0.367
Race-White	6.728 ± 3.425	2.519 ± 0.038
Race-Asian	17.79 ± 11.96	117.9 ± 0.045
Race-Hispanic	53.90 ± 19.00	9.669 ± 1.606
Income-White	132.2 ± 2.823	213.0 ± 0.190
Income-Black	108.9 ± 88.73	389.3 ± 0.026
Income-Indian	176.4 ± 7.351	700.9 ± 0.014
Income-Asian	17.76 ± 8.051	386.2 ± 0.077
Income-Other	12.63 ± 6.762	411.3 ± 0.136
Income-Hispanic	175.2 ± 4.667	404.7 ± 0.020
RMSE	0.627 ± 0.054	0.766 ± 0.036
Unfairness	0.050 ± 0.001	0.0024 ± 0.0001

Fair GP increases bandwiths wrt. GP, learned function varies less in those dimensions

Part III

Physics in Machine Learning models

Outline of part III

11. Motivation and problem statement

12. Results
    - Consistency with models for biophysical parameter estimation
    - **Consistency with ancillary in situ data**
    - Learning patterns of forced warming under uncertain climate variability

1/ Motivation

- Reconcile data-driven models with physics modeling by incorporating physical knowledge in Machine Learning models
- We consider algorithmic fairness (fairness by statistical parity)
- Encoding fairness and consistency with domain knowledge play similar roles

Learn $\hat{y} = f(x)$ such that $\|y - \hat{y}\|_2^2$ is minimized, and

reinforce $\quad \hat{y} \not\perp\!\!\!\perp s \quad$ or $\quad \hat{y} \perp\!\!\!\perp s,$

Consistency with models for biophysical parameter estimation

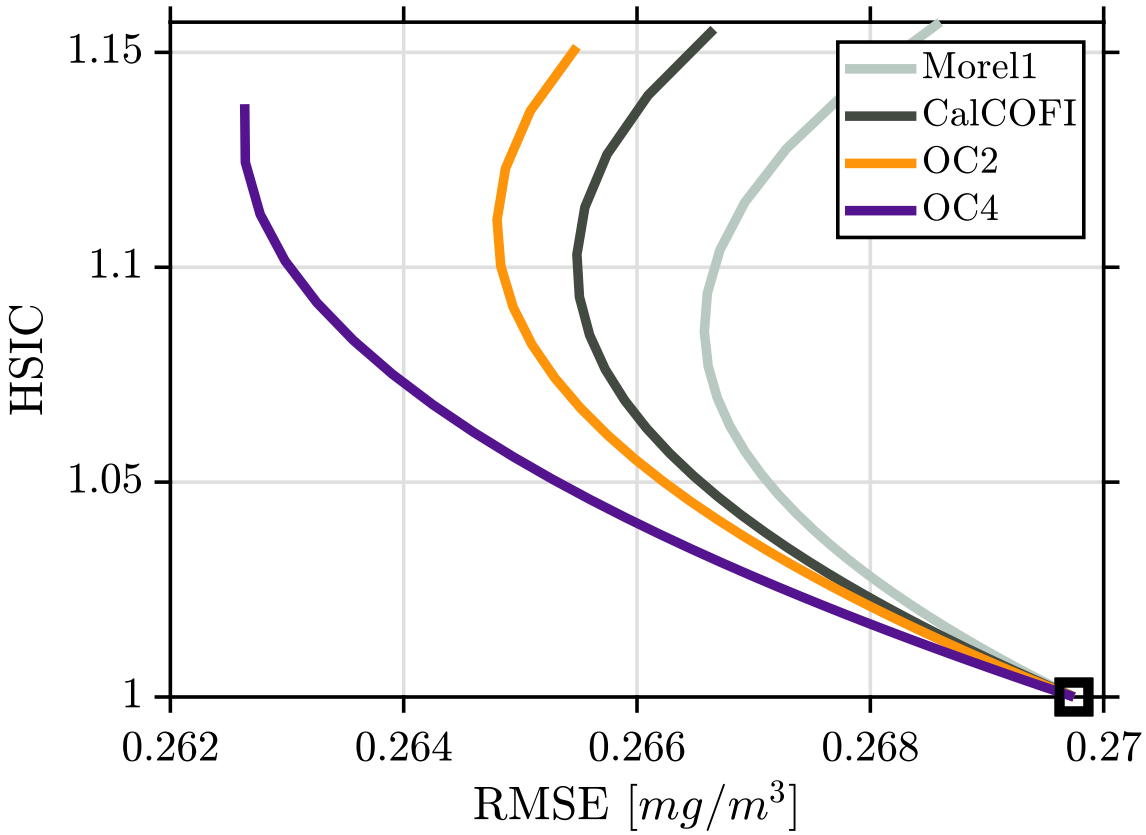
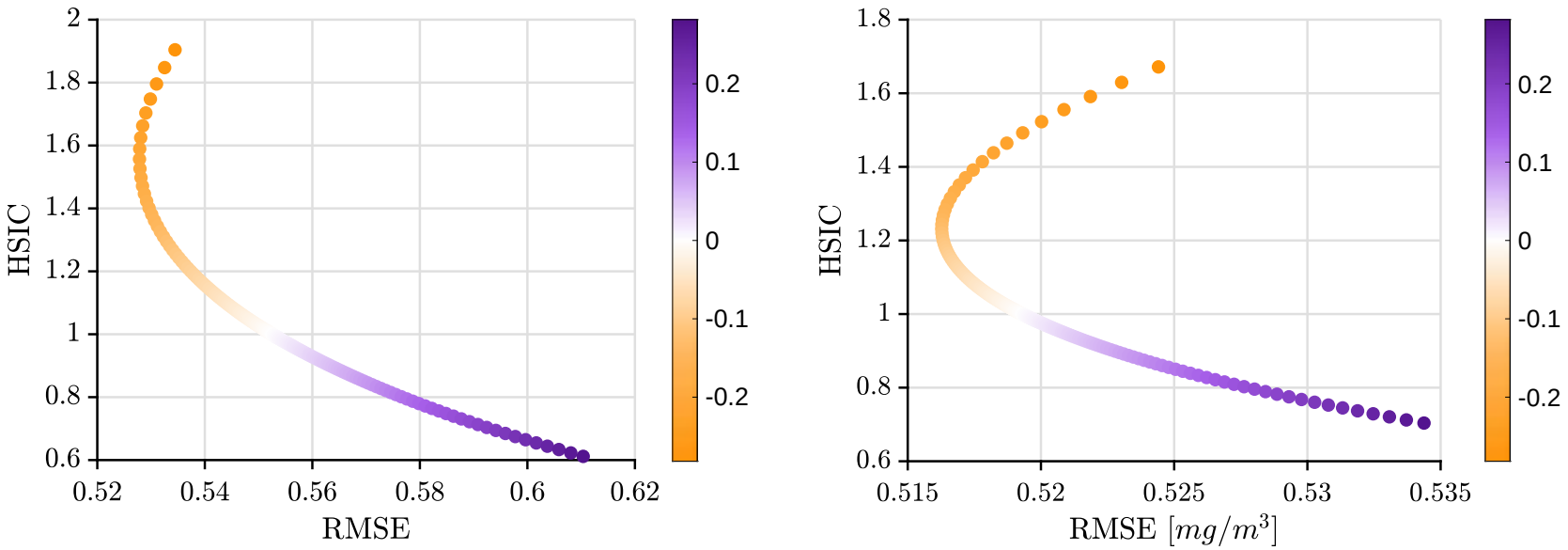


Figure: Consistency-vs-accuracy (HSIC-vs-RMSE) paths when including information from parametric models (Morel1, CalCOFI 2-band linear, OC2, and OC4) via a dependence-based regularizer in the PKL method. The black square corresponds to $\mu = 0$ while all the other points on the path correspond to increasing consistency for $\mu < 0$.

Consistency with ancillary in situ data



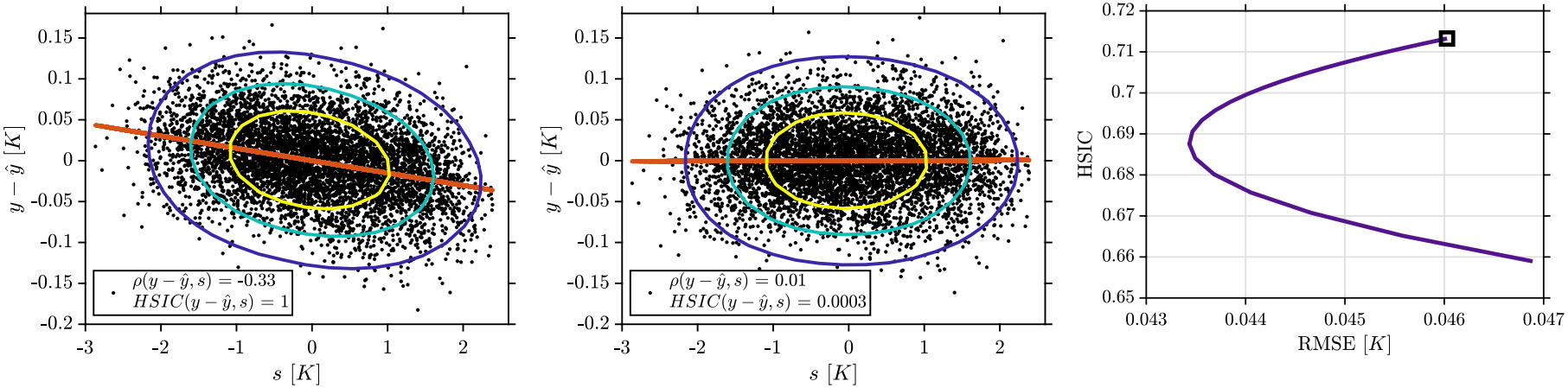
(a) LAI (y) and fCOVER (s)

(b) Chla (y) and LAI (s)

Figure: Consistency between a predicted variable and ancillary in situ measurements. Prediction of LAI with fCOVER as the ancillary variable (left), and prediction of Chla with LAI as the ancillary variable (right) for different values of μ indicated in the colorbar. In both cases, lower prediction RMSE can be obtained by encouraging consistency.

2/ Experiments

Learning patterns of forced warming under climate variability



(a) Standard regression

(b) Physics-aware regression

(c) Consistency-vs-accuracy path diagram

Figure: Scatter plot between residuals $r = y - \hat{y}$ and the Niño3.4 ancillary variable s (capturing the ENSO3.4 region variability) for the standard regression (a) and the physics-aware regression (b). We give both the Pearson's correlation coefficient between residuals $y - \hat{y}$ and s, and the achieved independence with HSIC between $y - \hat{y}$ and s. The (c) plot shows the path diagram for the most accurate model. When reducing dependence on the ENSO variability, the prediction error of the model decreases until a trade-off (turning point) appears between consistency and accuracy (c).

Thank you for your attention!

Questions?