Interpretable Machine Learning

Permutation Feature Importance (PFI)





Figure: Bike Sharing Dataset

PERMUTATION FEATURE IMPORTANCE (PFI) E Breiman (2001)

Idea: "Destroy" feat and interest x_j by perturbing it is it jit becomes uninformative, e.g. randomly permute lobshim x_j (marginal distribution $\mathbb{P}(x_j)$ is tays the same). PPF for features x_j using tested at a:D:D:

- ullet Measure the error without permuting feat and with permuted feat. values $ilde{x}_S$
- Repeate ermuting the feater (e.g., m times) and avgo the difference of both errors; both errors.

$$\begin{split} \widehat{\mathit{PFI}}_S &= \tfrac{1}{m} \sum_{R} \widehat{\mathit{PFI}}_S \mathcal{R}_{emp} \widehat{(\hat{f}, \mathcal{D})} \widehat{\mathcal{R}}_{emp} \widehat{(\hat{f}, \mathcal{D})} \mathcal{D} \mathcal{R}_{emp} \widehat{(\hat{f}, \mathcal{D})} \widehat{\mathcal{R}}_{emp} \widehat{(\hat{f}, \mathcal{D})} = \tfrac{1}{n} \sum_{(x,y) \in \mathcal{D}} L(\widehat{f}(x), y) \\ \text{where } \mathcal{R}_{emp} (\widehat{f}, \mathcal{D}) &= \tfrac{1}{n} \sum_{(x,y) \in \mathcal{D}} L(\widehat{f}(x), y) \end{split}$$

PERMUTATION FEATURE IMPORTANCE (PFI) (2001)

Idea:a"D estroy" feat-of interest x/by perturbing it sit alt becomes uninformative leight mative, e. a. randomly permute lobs bin x_i (marginal distribution $\mathbb{P}(x_i)$ istays the same), vs the same). PFFfor features x_s using test data \mathcal{D} : \mathcal{D} :

- Measure the error without permuting feat and with permuted feat, values \tilde{x}_S
- Repeat permuting the feat (e.g., m times) and avgo the difference of both errors both errors.

$$\widehat{\mathit{PFI}}_S = \tfrac{1}{m} \, \sum_{k=1}^m \tfrac{1}{m} (\hat{\Sigma}^m_{k=1} \, \mathcal{R}_{\mathsf{emp}} \widehat{(t_f}, \widetilde{\mathcal{D}}^{s_f}_{(k)}) \, \mathcal{D}) \mathcal{R}_{\mathsf{emp}} \widehat{(t_f}, \mathcal{D}), \\ \underset{\mathsf{emp}}{\mathsf{emp}} (\hat{t}, \mathcal{D}) = \tfrac{1}{n} \, \sum_{(x,y) \in \mathcal{D}} L(\hat{t}(x), y)$$

The data \mathcal{D} where x_s where x_s where x_s with x_s with x_s in x_s with x_s with x_s with x_s in x_s

The data \mathcal{D} where x_S is replaced with \tilde{x}^S is denoted as $\tilde{\mathcal{D}}^S$.

Example of permuting feature x_S with $S = \{1\}$ and m = 6:

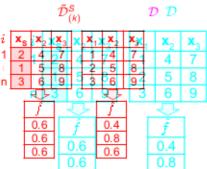
Note: The S in x_s refers to a Subset of features for which we are interested in their effect on the prediction

Here: We calculate the feature importance for one feature at a time |S| = 1.

			$\tilde{\mathcal{D}}$		\mathcal{D} \mathcal{D}				
i	xs	\mathbf{X}_2	X ₃	X	x ₁ × _x	2	g x ₁	X ₂	X ₃
1	2	14	27	4	1 74		7.1	4	7
1	1	5	8	5	2 35		34	5	8
n,	3	0	39	в	3 6	1	4	6	9

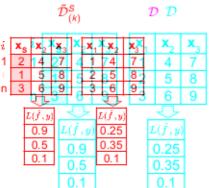


- **1. Perturbation:** Sample feature values from the distribution of $x_S(P(X_S))$.
 - 1. \Rightarrow Randomly permute feature x_s values from the distribution of x_s ($P(X_s)$). \Rightarrow Replace original feature with permuted feature x_s and create data $\tilde{\mathcal{D}}^s$
 - \Rightarrow Replace original feature with permuted feature \tilde{x}_S and create data \tilde{D}^S containing \tilde{x}_S original feature with permuted feature \tilde{x}_S and create data \tilde{D}^S containing \tilde{x}_S .





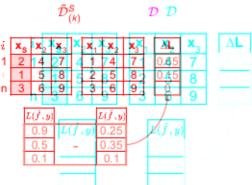
- 1. Perturbation: Sample feature values from the distribution of $x_S(P(X_S))$.
 - ⇒ Randomly permute feature x_S
 - 1 Preplace original feature with permuted feature \vec{x}_S and create data $(\vec{D}_S X_S)$). containing (\vec{x}_S) permute feature x_S
- 2. Predictions Make predictions for both data of e.g. and $\tilde{\mathcal{D}}^{snd}$ create data $\tilde{\mathcal{D}}^{S}$ containing \tilde{x}_{S}
- 2. Prediction: Make predictions for both data, i.e., \mathcal{D} and $\tilde{\mathcal{D}}^S$





3. Aggregation:

- 3. A Compute the loss for each observation in both data sets
 - Compute the loss for each observation in both data sets





3. Aggregation:

- 3. A Compute the loss for each observation in both data sets
 - Take the difference of both dosses ΔL for each observations
 - Take the difference of both losses ΔL for each observation

$$\mathcal{R}^{s}_{\mathsf{emp}}(\hat{f}(\tilde{\mathcal{D}}^{s}_{(k)}) \overset{.}{\to} \mathcal{R}_{\mathsf{emp}}(\hat{f}_{\mathfrak{p}}(\hat{\mathcal{D}}) \, \mathcal{D})$$

				_					
i Xs	X,	\mathbf{x}_3	X	х,	X.	X	ΔŁ	\mathbf{x}_3	ΔL
1 2	4	27	4	1	74	7	0.65	7	0.65
: 1	5	18	5	2	35	82	0.15	8	0.2675
n 3	6	9	ā	3	6	9	0	a	0

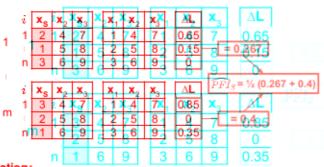


3. Aggregation:

- 3. Appendix the loss for each observation in both data sets
 - Take the difference of both losses ΔL for each observation
 - Take the difference of both losses A/ for each observation
 Average this change in loss across all observations
 - Note: This is equivalent to computing \mathcal{R}_{emp} on both data sets and taking the difference the difference on both data sets and taking the difference

Interpretable Machine Learning - 2 / 12

$$\mathcal{R}^{s}_{\mathsf{emp}}(\hat{f}(\widetilde{\mathcal{D}}^{s}_{(k)}) \to \mathcal{R}^{s}_{\mathsf{emp}}(\hat{f}_{\mathfrak{p}}(\widehat{\mathcal{D}}) \, \mathcal{D})$$

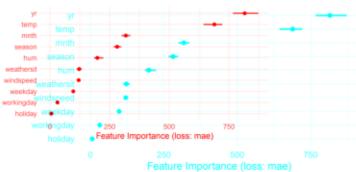




3. Aggregation:

- 3. A Compute the loss for each observation in both data sets
 - Take the difference of both losses ΔL for each observations
 - Average this change in loss across all observations bservation
 - Repeat perturbation and average over multiple repetitions
 - Repeat pertubation and average over multiple repetitions

EXAMPLE: BIKE SHARING DATASET





Interpretation:

- Year (yr) and Temperature (temp) are most important features
 - Destroying information about yr by permuting it increases mean absolute error of model by 816 Temperature (temp) are most important features
 - 5% and 95% guartifet of repetitions due multiplie permutations are shown als te error of model by 816
 - error bars 95% quantile of repetitions due multiple permutations are shown as error bars

COMMENTS ON PFI

Interpretation: PFIfis the line rease of model renormation is destroyed

COMMENTS ON PFI

Interpretation: PFI is the increase of model error when feature's information as on is destroyed

- Results can be unreliable due to random permutations
- Results can be unreliable due to random permutations is
 - ⇒ Solution: Average results over multiple repetitions



COMMENTS ON PFF

Interpretation: PFIlis the increase of model enough when feature's information is destroyed

- Results can be unreliable due to random permutations
- Results can be unreliable due to random permutations is
 - ⇒ Solution: Average results over multiple repetitions
 - Permuting features despite correlation with other features can lead to unrealistic combinations
- Permuting features despite correlation with other features can lead to unrealistic xtrapolation issue combinations of feature values (since under dependence $\mathbb{P}(x_i, x_{-i}) \neq \mathbb{P}(x_i)\mathbb{P}(x_{-i})$) \leadsto Extrapolation issue

COMMENTS ON PFF

- Interpretation: PFI is the increase of model error when feature's information as on is destroyed
- Results can be unreliable due to random permutations
- Results can be unreliable due to random permutations is
 - ⇒ Solution: Average results over multiple repetitions
- Permuting features despite correlation with other features can lead to unrealistic combinations
- Permuting features despite correlation with other features can lead to unrealistic xtrapolation issue combinations of feature values (since under dependence
 - P(x̄, xū) ≠ 4P(x̄,)P(xū)) P(xū)) es iExtrapolation lissure action effects with other features

 ⇒ Permutation also destroys information of interactions where permuted feature is involved
- PEI automatically includes importance of interaction effects with other features in PEI score
 - ⇒ Permutation also destroys information of interactions where permuted feature is involved
 - ⇒ Importance of all interactions with the permuted feature are contained in PFI score

COMMENTS ON PFF

Interpretation: PFI is the increase of model error when feature's information as on is destroyed

- Results can be unreliable due to random permutations
- Results can be unreliable due to random permutations is
 - ⇒ Solution: Average results over multiple repetitions
- Permuting features despite correlation with other features can lead to unrealistic combinations
- Permuting features despite correlation with other features can lead to unrealistic xtrapolation issue combinations of feature values (since under dependence
- P(x/xy)P(xy)P(xy)P(xy)) See Extrapolation listing action effects with other features

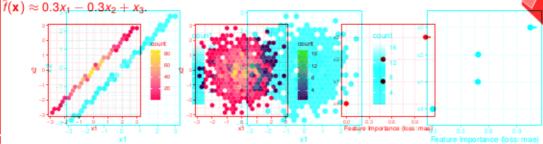
 ⇒ Permutation also destroys information of interactions where permuted feature is involved
- PEL automatically includes importance of interaction effects with other features in PEL score
 - ⇒ Permutation also destroys information of interactions where permuted feature
 - ois involvedation of PFI depends on whether training or test data is used
 - \Rightarrow Importance of all interactions with the permuted feature are contained in PFI score
- . Interpretation of PFI depends on whether training or test data is used

COMMENTS ON PFI - EXTRAPOLATION

Example t:Let $y \neq x_3 + \epsilon_y$ with $\epsilon_y \wedge (0,0.1)$ where $x_1 r = x_1 + \epsilon_y = x_1 + \epsilon_y$ are ϵ_2 are highly correlated highly correlated $(\epsilon_1 \sim N(0,1),\epsilon_2)$ and (0,0.01), and (0,0.01), and (0,0.01), and (0,0.01). All noise terms are ϵ_3 inches N(0,0.01). Fall noise terms are ϵ_3 inches N(0,0.01). Fall noise terms are N(0,0.01).

COMMENTS ON PFI - EXTRAPOLATION

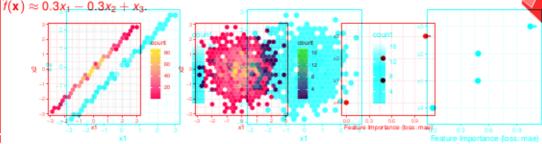
Example: Let $y \neq x_3 + \epsilon_y$ with (0,0.0.1) where $(x_1 r = x_1 + \epsilon_y) = x_1 + \epsilon_y$ are $(x_1 r = \epsilon_y) = x_1 + \epsilon_y$ are $(x_1 r = \epsilon_y) = x_1 + \epsilon_y$ are $(x_1 r = \epsilon_y) = x_1 + \epsilon_y$ are $(x_1 r = \epsilon_y) = x_1 + \epsilon_y$ are $(x_1 r = \epsilon_y) = x_1 + \epsilon_y$ are $(x_1 r = \epsilon_y) = x_1 + \epsilon_y$ are $(x_1 r = \epsilon_y) = x_1 + \epsilon_y$ are $(x_1 r = \epsilon_y) = x_1 + \epsilon_y$ are $(x_1 r = \epsilon_y) = x_1 + \epsilon_y$ are $(x_1 r = \epsilon_y) = x_1 + \epsilon_y$ are $(x_1 r = \epsilon_y) = x_1 + \epsilon_y$ are $(x_1 r = \epsilon_y) = x_1 + \epsilon_y$ are $(x_1 r = \epsilon_y) = x_1 + \epsilon_y$ are $(x_1 r = \epsilon_y) = x_1 + \epsilon_y$ are $(x_1 r = \epsilon_y) = x_1 + \epsilon_y$ are $(x_1 r = \epsilon_y) = x_1 + \epsilon_y$ are $(x_1 r = \epsilon_y) = x_1 + \epsilon_y$ are $(x_1 r = \epsilon_y) = x_1 + \epsilon_y$ and $(x_1 r = \epsilon_y) = x_1 + \epsilon_y$ are $(x_1 r = \epsilon_y) = x_1 + \epsilon_y$ and $(x_1 r = \epsilon_y) = x_1 + \epsilon_y$ are $(x_1 r = \epsilon_y) = x_1 + \epsilon_y$ and $(x_1 r = \epsilon_y) = x_1 + \epsilon_y$ are $(x_1 r = \epsilon_y) = x_1 + \epsilon_y$ and $(x_1 r = \epsilon_y) = x_1 + \epsilon_y$ are $(x_1 r = \epsilon_y) = x_1 + \epsilon_y$ and $(x_1 r = \epsilon_y) = x_1 + \epsilon_y$ and $(x_1 r = \epsilon_y) = x_1 + \epsilon_y$ are $(x_1 r = \epsilon_y) = x_1 + \epsilon_y$ and



Hexbin plot of x_1, x_2 before permuting x_1 (left), after permuting x_1 (center), and PFI scores (right) of of x_1, x_2 before permuting x_1 (left), after permuting x_1 (center), and PFI scores (right)

COMMENTS ON PFI - EXTRAPOLATION

Example be Lettey $\neq x_3 + \epsilon_y$ with $d_y = N(0,0,0,0)$ where $|x_1| = \epsilon_1 + \epsilon_2 = \epsilon_2$ are highly correlated highly correlated $|x_1| = \epsilon_1 + \epsilon_2 = \epsilon_2$ are highly correlated highly correlated $|x_1| = \epsilon_1 + \epsilon_2 = \epsilon_2$ are highly correlated highly correlated $|x_1| = \epsilon_1 + \epsilon_2 = \epsilon_2$ are highly correlated highly



Hexbin plot of x_1, x_2 before permuting x_1 (left), after permuting x_1 (center), and PFI scores (right) $x_1 = x_1 + x_2 + x_3 + x_4 + x_4 + x_5 +$

 $\{\mathbf{x} \geq \mathbb{P}(\mathbf{x}) \leq \mathbf{0}\}$ as $0.3x_1 = 0.3x_2 \approx 0$ or the prediction $\hat{f}(\mathbf{x})$ for $\{\mathbf{x} \geq \mathbb{P}(\mathbf{x}) > 0\}$ as $0.3x_1 = 0.3x_2 \approx 0$ $\Rightarrow \mathsf{PFF}$ evaluates model on unrealistic observables $\mathbb{P}(\mathbf{x}) \neq x_1, x_2$ are considered even relevant (PFI > 0) relevant (PFI > 0)

Interpretable Machine Learning - 5 / 12

COMMENTS ON PFI - INTERACTIONS

Example: Let ex₁ x₁ ... , x₄ be independently rand uniformly rsampled from {e-r1, 1} and } and

$$y := x_1 x_2 + x_3 + \epsilon_Y \text{ with}_3 \epsilon_Y \sim_Y N(0, 1)_Y \sim N(0, 1)$$

Fitting a $\pm M$ yields $\hat{f}(\hat{x}) \approx x_1 x_2 \pm x_3 \cdot x_3$.



COMMENTS ON PFF-INTERACTIONS

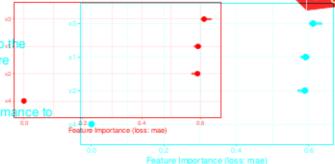
Example: Let ext, x4,..., x4 be independently rand uniformly sampled from {e-r1, 1} and } and

$$y := x_1 x_2 + x_3 + \epsilon_Y \text{ with } \epsilon_Y - \epsilon_Y N(0, 1)$$

Fitting a LM yields $\hat{f}(\hat{x}) \approx x_1 x_2 + x_3 \cdot x_3$.

Although | |x3 alone contributes as much to to the the prediction as | |x4 and | |x2 jointly all three are are considered equally relevant.

⇒ PFI does not fairly attribute the performance to performance to the individual features.



Interpretable Machine Learning - 6 / 12

ICOMMENTS ON PELL TEST VS. TRAINING DATA

Example Lex₁ y_1 ..., x_{20} y_2 are independently sampled from $\mathcal{U}(n-3.0,-10)$. An x globost boost model with model with default hyperparameters as fit on a small training set of 50 observations model overfits heavily. The model overfits heavily.

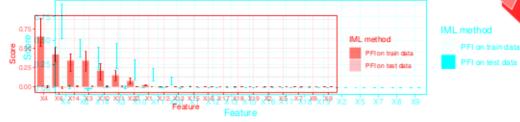


Figure: While PELon test data considers all features to be irrelevant. PELon train data exposes the data exposes the features on which the model overfitted.

COMMENTS ON PFIL TEST VS. TRAINING DATA

Example $1, x_1, \dots, x_{20}$ y, are independently sampled from $\mathcal{U}(n-30, -10)$. An xg-boost boost model with model with default hyperparameters as fit on a small training set of 50 observations model overfits heavily. The model overfits heavily.

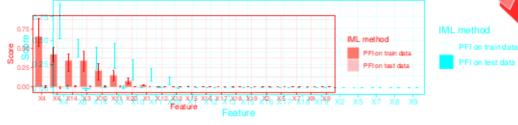


Figure: While PELon test data considers all features to be irrelevant. PELon train data exposes the features on which the model overfitted.

Why? PFI can only be nonzero if the permutation breaks a dependence in the data. Spurious Spurious correlations help the model perform well on train data but are not present in the test data. Spurious help the model perform well on train data, but are not present in the test data. Spurious the test data. Spurious the test data in the test data. Spurious the test data in the test data in the test data. Spurious the test data in the test data in the test data in the test data. Spurious the test data in the test data in the test data in the test data. Spurious the test data in the test data in the test data in the test data in the test data.

⇒ If you are interested in which features help the model to generalize, apply PFF on test data.

Interpretable Machine Learning - 7 / 12

IMPLICATIONS OF PFI

Can we get in sight into whether the the ...

- feature x_j is icausal for the prediction 2n?
 - PFI_FI≠ 0 ⇒ model reliesion x_j x_j
 - As the training vs/test data example demonstrates, the donverse does not s not hold hold



IMPLICATIONS OF PFI

Can we get in sight into whether the the ...

- feature x_j is icausal for the prediction? n?
 - PFI_i i≠ 0 ⇔ modelireliesion x_i x_i
 - As the training vs/ test data example demonstrates, the converse does not s not hold
- \bullet feat $b P d x_j$ contains prediction-relevant information?
- **9** feature x_i contains prediction relevant information? variates x_{-i} or both (due to extrapolation)
 - • PFIj \Rightarrow 0 t \Rightarrow xylist dependent of (yeogat's covariates xhg on boths (due for y or not) \Rightarrow PFIj = 0 extrapolation)
 - x_j is not exploited by model (regardless of whether it is useful for y or not)
 ⇒ PFI_i = 0



IMPLICATIONS OF PFI

Can we get in sight into whether thethe ...

- feature x_j is causal for the prediction? n?
 - ••PFI $_i \neq 0 \Leftrightarrow$ modelirelies on $x_i \times x_i$
 - As the training vs/ test data example demonstrates, the converse does not s not hold
- \bullet feat $b P d x_j$ contains prediction-relevant information?
- **6** feature x contains prediction relevant information? variates x_{-i} or both (due to extrapolation)
 - • $PFI_j \Rightarrow 0$ t $\Rightarrow x_0$ is the dependent of (year at slow a real sector y or y o
- \bullet modextrapplation ccess to x_i to achieve it's prediction performance?
 - x_i is not exploited by model (regardless of whether it is useful for y or not) ⇒ PFI_i = 0
- \odot model requires access to x_i to achieve it's prediction performance?
 - · As the extrapolation example demonstrates, such insight is not possible

TESTING IMPORTANCE (PIMP) Altmann et al. (2010)

PIMP was originally lintroduced for frandom for est/s built-in permutation feature importance

TESTING IMPORTANCE (PIMP) Altmann et al. (2010)

- PIMP was originally lintroduced for frandom for est's built-in permutation feature ature importance.
- eimportance estigates whether the PFI score significantly differs from 0
- PIMR investigates whether the RFh score significantly differs from 0
 - ⇒ Useful because PFI can be non-zero due to stochasticity

TESTING IMPORTANCE (PIMP) (2 Altmann et al. (2010))

- PIMP was originally lintroduced for frandom forest's built-in permutation feature at ure importance importance eimportance estimates whether the PFI score significantly differs from 0
- PIMR investigates whether the IRFh score significantly differs from 0
 Useful because IPFI can be non-zero due to stochasticity (the target y (unimportant))
- PIMP tests the H₀-hypothesis: Feature is independent of the target y (unimportant)

TESTING IMPORTANCE (PIMP) (2 Altmann et al. (2010))

PIMP was originally lintroduced for frandom forest's built-in permutation feature at ure importance importance eimportance estimates whether the PFI score significantly differs from 0

- PIMR investigates whether the RFh score significantly differs from 0
- > Useful because PFI can be non-zero due to stochasticity f the target y (unimportant)
- PIMP tests the Ho-hypothesis: Feature is independent of the target VFI scores (repeat)
 (unimportant) y breaks relationship to all features
- Sampling under H₀: Rel mute: target iy, retrain model; compute PFI scoreses under H₀ (repeat)
 - ⇒ Permuting y breaks relationship to all features
 - ⇒ By computing PFI scores again, we obtain distribution of PFI scores under H₀

TESTING IMPORTANCE (PIMP) (2 Altmann et al. (2010))

- PIMP was originally lintroduced for frandom forest's built-in permutation feature at ure importance importance eimportance estimates whether the PFI score significantly differs from 0
- PIMP investigates whether the PFh score significantly differs from 0
- > Useful because PFI can be non-zero due to stochasticity f the target y (unimportant)
- PIMP tests the do-hypothesis: Feature is independent of the target VFI scores (repeat)
 (unimportant) g y breaks relationship to all features
- Sampling under H₀: Permute target iy, retrain models compute PFI scoreses under H₀
 - (repeat) ute p-value the tail probability under H_0 and use it as a new importance measure
 - ⇒ Permuting y breaks relationship to all features
 - \Rightarrow By computing PFI scores again, we obtain distribution of PFI scores under H_0
- Compute p-value the tail probability under H₀ and use it as a new importance measure

TESTING IMPORTANCE (PIMP)

PIMP algorithm:::

- Formn∈ {1,1,..,n_{repetitions}}:s}:
 - Permute response vector yr y
 - Retrain model with data X and permuted yd y
 - Compute feature importance RFIP for each feature ji (under H₀): H₀)



TESTING IMPORTANCE (PIMP)

PIMP algorithm:::

- Form∈ {1,1,.,n_{repetitions}}:s}:
 - Permute response vector yr y
 - Retrain model with data X and permuted yd y
 - Compute feature importance RFIT for each feature ji (under H₀): H₀)
- Train model with X and unpermuted yd y



TESTING IMPORTANCE (PIMP))

PIMP algorithm:::

- Form∈ {1{1, ..., n_{repetitions}}:s}:
 - Permute response vector y
 - Retrain model with data X and permuted yd y
 - Compute feature importance RFITTor each feature ji (under H₀): H₀)
- Train model with X and unpermuted yd y
- Foreach feature je €1,(1,..,p);p}:
 - Fit probability idistribution of the lifeature importance values $PFI_j^m P_i FI_j^m$, $m \in \{1, \ldots, n_{repetitions}\}$ $m \in \{1, \ldots, n_{repetitions}\}$ (choice between Gaussian, lognormal gamma) or non-parametric) are importance PFI_i for the model without permutation of y (under H_1)
 - Compute feature importance PFI_j for the model without permutation of y (under H₁)
 - Retrieve the p-value of PFI; based on the fitted distribution



PIMP FOR EXTRAPOLATION EXAMPLE

Recall: $|y| = x_3 x_1 + \epsilon_y$ with $|x| > N(0,001,0,x_1)$, x_2 highly obrretated but independent of ent of y, x_4 is y in x_1 is independent of y, and all other variables: if itting a LMI yields, $y = 0.3x_1 - 0.3x_2 + x_3$. If $y = 0.3x_1 -$

- Histograms: H₀ distribution of PFI scores after permuting y (1000 repetitions)
- Histograms: H₁ distribution of PFI scores after permuting y (1000 repetitions)
 Red: PFI score estimated on unpermuted y (under H₁) → compare against H₀
- distribution score estimated on unpermuted y (under H_1) \rightarrow compare against H_0 distribution
- Results! Although PFI for x; and x2 is nonzero (red), PIMP considers them not significantly significantly relevant (p-value > 0.05)

DIGRESSION: MULTIPLE TESTING PROBLEM (* Romano et al. (2010)

- When should we reject the H₀-hypothesis for a feature?
- When should we reject the H₀-hypothesis for a feature?



DIGRESSION: MULTIPLE TESTING PROBLEM

- When should we reject the H₀-hypothesis for a feature?
- When should we reject the Heavypothesis for a feature? eed to be performed by PIMP
- The larger the number of features, the more tests need to be performed by PIMPt, the probab ~# Multiple testing problem: If multiplicity of tests (is not taken into account, they be large probability that some of the true H_0 -hypothesis is rejected (type-I error) by chance may be large

DIGRESSION: MULTIPLE TESTING PROBLEM (* Romano et al. (2010)

➤ Romano et al. (2010)

- When should we reject the H₀-hypothesis for a feature?
- When should we reject the Heavy pothesis for a feature? eed to be performed by PIMP
- The larger the number of teatures, the more tests need to be performed by PIMPit, the probability of tests is not taken into account, they be large probability that some of the true Harbypothesis is rejected (type-1 error) by olling an appropriate error chance may be larger wise error rate (FWE: probability of at least one type-I error)
- Accounting for multiplicity of individual tests can be achieved by controlling an appropriate error rate, e.g., the family-wise error rate (FWE: probability of at least one type-I error)

DIGRESSION: MULTIPLE TESTING PROBLEM

► Romano et al. (2010)

- When should we reject the H₀-hypothesis for a feature?
- When should we reject the Heavy pothesis for a feature? eed to be performed by PIMP
- The larger the number of features (the more tests need to be performed by PIMP), the probabilities ~# Multiple testing problem: If multiplicity of tests (is not taken into account, they be large
- probability, that some of the true Hothypothesis is rejected (type-Lerror) by olling an appropriate error chance may be largely-wise error rate (FWE: probability of at least one type-I error)
- Accounting for multiplicity of individual tests can be achieved by controlling anyhich rejects a null appropriate error rate, e.g., the family-wise error rate (FWE: probability of at formed parallel tests least one type-I error)
- One classical method to control the FWE is the Bonferroni correction which rejects a null hypothesis if its p-value is smaller than α/m with m as the number of performed parallel tests