A New Generalized Error Path Algorithm for Model Selection

Bin Gu[1,2] Charles X. Ling[1]

[1]Department of Computer Science
University of Western Ontario, Canada

[2]School of Computer & Software
Nanjing University of Information Science & Technology, China

July 9, 2015

Outline

- Background
- Generalized Error Path
- Experiments
- Conclusion

Cross Validation

Model selection with cross validation (CV) is very popular in machine learning.

- K-fold

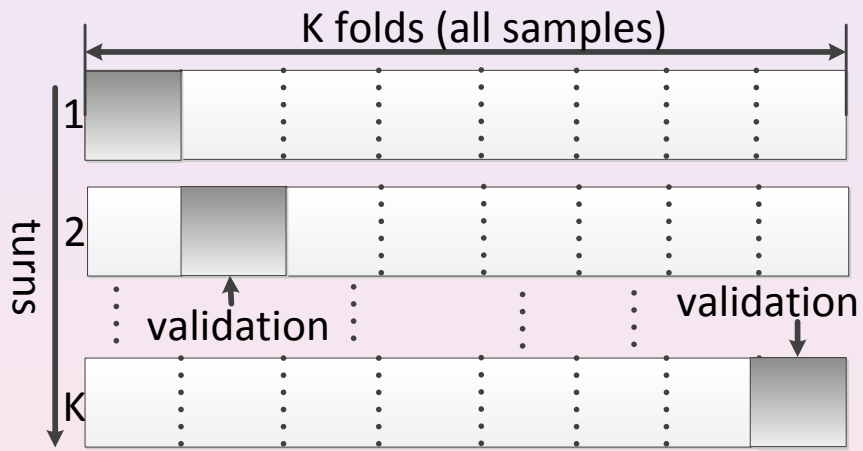


Figure: The schematic diagram of K-fold cross validation.

- Leave-one-out
- Repeated random sub-sampling procedures

Cross Validation

Model selection with cross validation (CV) is very popular in machine learning.

- K-fold

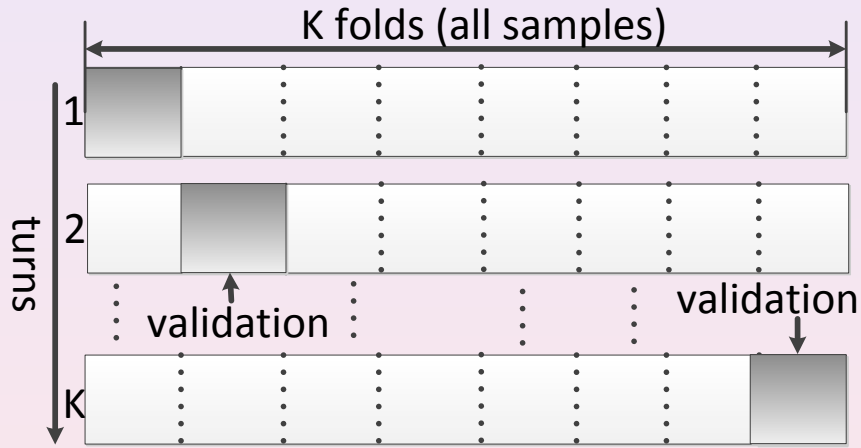


Figure: The schematic diagram of K-fold cross validation.

- Leave-one-out
- Repeated random sub-sampling procedures

Searching Strategies in CV

CV works with a group of candidate values of parameter

- The most popular strategy is grid searching
  - Grid search: the regularization parameter C of SVM is searched on a 18 grid linearly spaced in the region $\{(\log_2 C) | -9 \leq \log_2 C \leq 8\}$, as used in [Yang & Ong(2011)]
- Other searching strategies
  - Manual search
  - Random search

Searching Strategies in CV

CV works with a group of candidate values of parameter

- The most popular strategy is grid searching
  - Grid search: the regularization parameter C of SVM is searched on a 18 grid linearly spaced in the region $\{(\log_2 C) | -9 \leq \log_2 C \leq 8\}$, as used in [Yang & Ong(2011)]
- Other searching strategies
  - Manual search
  - Random search

Weakness of grid, manual, and random search strategies

- Only considering finite candidates due to the limited computing resources
- The ultimate goal of model selection is to find the model with the minimum CV error in the whole parameter space

Weakness of grid, manual, and random search strategies

- Only considering finite candidates due to the limited computing resources
- The ultimate goal of model selection is to find the model with the minimum CV error in the whole parameter space

Solution path

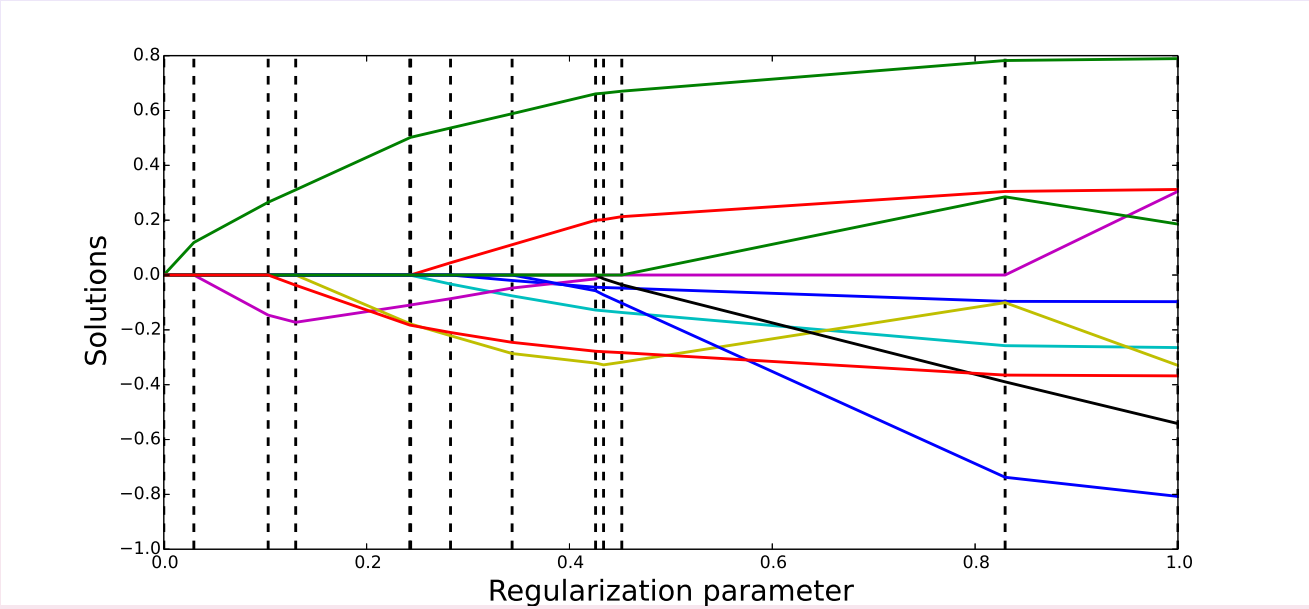


Figure: The solutions with respect to the regularization parameter.

- A finite number of representative solutions fit the entire solutions.

Solution path

Table: Representative solution path algorithms.

Problem	Task	Reference	Parameter	Exact
C-SVC	BC	[Hastie et al.(2004)]	Regularization parameter C	Yes
$2C$-SVC	BC	[Bach et al.(2006)]	Regularization parameters C_+, C_-	Yes
ε-SVR	R	[Gunter & Zhu(2007)]	Regularization parameter	Yes
ε-SVR	R	[Wang et al.(2008)]	Regularization parameter and ε	Yes
Lasso	R	[Rosset & Zhu(2007)]	Regularization parameter	Yes
KQR	R	[Takeuchi et al.(2009)]	Quantile order $\tau \in (0, 1)$	Yes
C-SVC	BC	[Ong et al.(2010)]	Regularization parameter C	Yes
C-SVC	BC	[Karasuyama et al.(2011)]	Regularization parameter C	No
ν-SVC	BC	[Gu et al.(2012)]	Regularization parameter ν	Yes
General	BC+R	[Giesen et al.(2012)]	Regularization parameter	No

Error path

- However, solution paths do not guarantee to find the model with minimum CV error.
- The only work on CV with global search is [Yang & Ong(2011)]
  - based on the solution paths of [Hastie et al.(2004), Ong et al.(2010)]
  - C-SVC
  - standard error function of binary classification
- The method is limited to the two solution path algorithms and error functions of binary classification

Error path

- However, solution paths do not guarantee to find the model with minimum CV error.
- The only work on CV with global search is [Yang & Ong(2011)]
  - based on the solution paths of [Hastie et al.(2004), Ong et al.(2010)]
  - C-SVC
  - standard error function of binary classification
- The method is limited to the two solution path algorithms and error functions of binary classification

Motivation

Design a generalized error path algorithm (GEP)

- incorporating more general error (or loss) functions
- working with more general solution path algorithms.
- obtaining minimum CV errors directly for a large variety of error (or loss) functions and solution path algorithms

Generalized Error Path

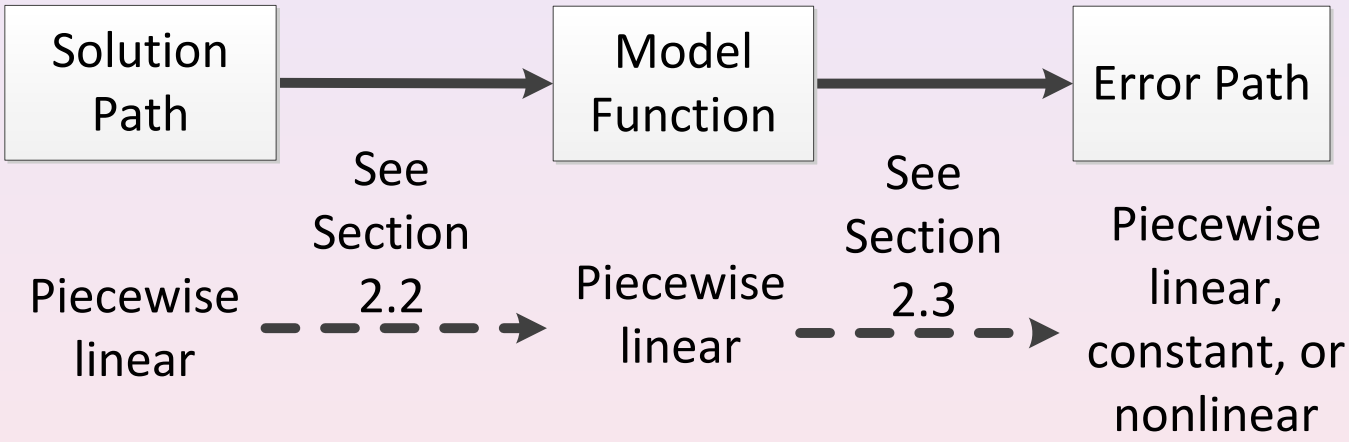
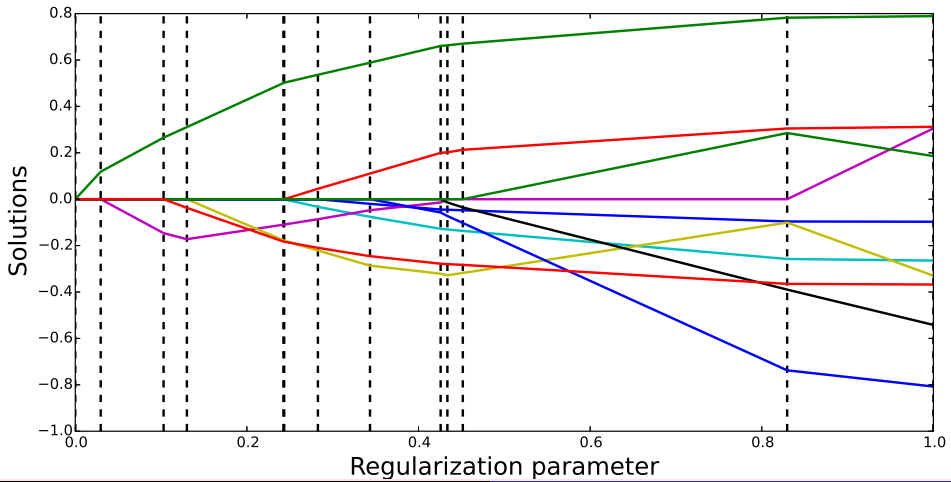


Figure: Model function builds a bridge from solution path to error path.

The piecewise linearity of the solution path

Definition

Suppose $\tilde{\alpha}(\lambda)$ is returned by a solution path. The solution $\tilde{\alpha}(\lambda)$ is called piecewise linear as a function of λ, if existing $a = a_0 < a_1 < a_2 < \cdots < a_m = b$, and the corresponding vectors $\beta^{[1]}, \beta^{[2]}, \ldots, \beta^{[m]}$, such that the solution $\hat{\alpha}(\lambda)$ is given exactly or approximately, by $\tilde{\alpha}(a_{k-1}) + \beta^{[k]}(\lambda - a_{k-1})$, $\forall \lambda \in [a_{k-1}, a_k]$.



The piecewise linearity of the solution path

Table: Representative solution path algorithms.

Problem	Task	Reference	Parameter	Exact	Piecewise
C-SVC	BC	[Hastie et al.(2004)]	Regularization parameter C	Yes	Linear
$2C$-SVC	BC	[Bach et al.(2006)]	Regularization parameters C_+, C_-	Yes	Linear
ε-SVR	R	[Gunter & Zhu(2007)]	Regularization parameter	Yes	Linear
ε-SVR	R	[Wang et al.(2008)]	Regularization parameter and ε	Yes	Linear
Lasso	R	[Rosset & Zhu(2007)]	Regularization parameter	Yes	Linear
KQR	R	[Takeuchi et al.(2009)]	Quantile order $\tau \in (0, 1)$	Yes	Linear
C-SVC	BC	[Ong et al.(2010)]	Regularization parameter C	Yes	Linear
C-SVC	BC	[Karasuyama et al.(2011)]	Regularization parameter C	No	Linear
ν-SVC	BC	[Gu et al.(2012)]	Regularization parameter ν	Yes	Linear
General	BC+R	[Giesen et al.(2012)]	Regularization parameter	No	Linear

From Solution Path to Model Function

A model with a linear representation $f(x) = \langle G(x), \alpha \rangle$

1. For C-SVC, $2C$-SVC, ν-SVC, the models are $f(x) = \sum_i^l \alpha_i y_i K(x, x_i) + \alpha_0$.
2. For ε-SVR, the model is $f(x) = \sum_i^{2l} \alpha_i z_i K(x, x_i) + \alpha_0$.
3. For Lasso, the model is $f(x) = x^T \alpha$.
4. For KQR, the model is $f(x) = \sum_i^l \alpha_i K(x, x_i) + \alpha_0$.

From Solution Path to Model Function

Lemma

Given an interval $[a_{k-1}, a_k]$ *in a solution path, and the corresponding vector* $\beta^{[k]}$*. If the model is with linear representation, there exists a scale* $\gamma^{[k]}$ *such that the model function* $f_\lambda(x)$ *can be represented as*
$f_\lambda(x) = f_{a_{k-1}}(x) + \gamma^{[k]}(\lambda - a_{k-1}), \forall \lambda \in [a_{k-1}, a_k]$.

1. For C-SVC, $2C$-SVC, and ν-SVC, $\gamma = \sum_i^l \beta_i y_i K(x, x_i) + \beta_0$.
2. For ε-SVR, $\gamma = \sum_i^{2l} \beta_i z_i K(x, x_i) + \beta_0$.
3. For Lasso, $\gamma = x^T \beta$.
4. For KQR, $\gamma = \sum_i^l \beta_i K(x, x_i) + \beta_0$.

From Solution Path to Model Function

Lemma

Given an interval $[a_{k-1}, a_k]$ *in a solution path, and the corresponding vector* $\beta^{[k]}$*. If the model is with linear representation, there exists a scale* $\gamma^{[k]}$ *such that the model function* $f_\lambda(x)$ *can be represented as*
$f_\lambda(x) = f_{a_{k-1}}(x) + \gamma^{[k]}(\lambda - a_{k-1}), \forall \lambda \in [a_{k-1}, a_k]$.

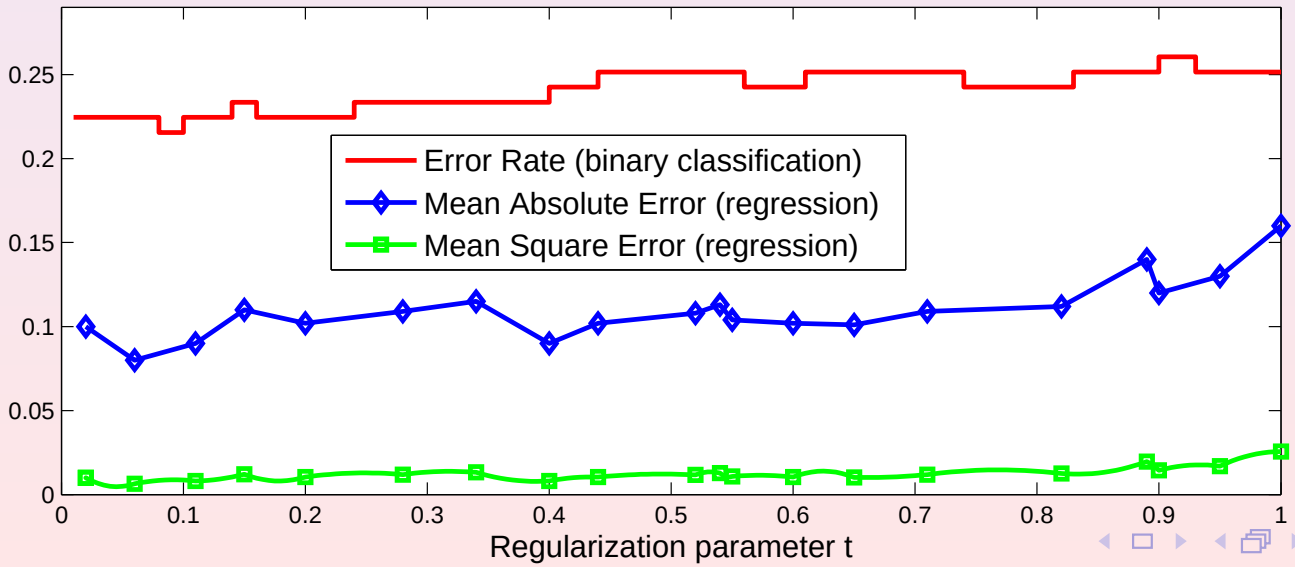
1. For C-SVC, $2C$-SVC, and ν-SVC, $\gamma = \sum_i^l \beta_i y_i K(x, x_i) + \beta_0$.
2. For ε-SVR, $\gamma = \sum_i^{2l} \beta_i z_i K(x, x_i) + \beta_0$.
3. For Lasso, $\gamma = x^T \beta$.
4. For KQR, $\gamma = \sum_i^l \beta_i K(x, x_i) + \beta_0$.

From Model Function to Error Path

The error path on the entire interval $[a, b]$ could be

1. piecewise quadratic,
2. piecewise linear,
3. or piecewise constant

w.r.t. λ, on common error (loss) functions for regression and binary classification problems.



Piecewise Quadratic

Square error (loss) function $L(\widetilde{y}, \widehat{y}) = (\widetilde{y} - \widehat{y})^2$ for regression

$$E(\hat{\alpha}(\lambda), \mathcal{V}, L) = \frac{1}{\ell} \sum_{i=1}^{\ell} L(\widetilde{y}_i, \widehat{y}_i) \tag{1}$$

$$= \frac{1}{\ell} \sum_{i=1}^{\ell} (\widetilde{y}_i - f_\lambda(\widetilde{x}_i))^2$$

$$= \frac{1}{\ell} \sum_{i=1}^{\ell} (\gamma_i^{[k]})^2 \lambda^2 - \frac{1}{\ell} \sum_{i=1}^{\ell} 2 \times \gamma_i^{[k]} (\widetilde{y}_i - f_{a_{k-1}}(\widetilde{x}_i) + \gamma_i^{[k]} a_{k-1}) \lambda$$

$$+ \frac{1}{\ell} \sum_{i=1}^{\ell} \left(\widetilde{y}_i - f_{a_{k-1}}(\widetilde{x}_i) + \gamma_i^{[k]} a_{k-1} \right)^2$$

The mean square error (MSE) is piecewise quadratic w.r.t. λ, in the whole interval $[a, b]$

Piecewise Linear

The absolute error (loss) function $L(\widetilde{y}, \widehat{y}) = |\widetilde{y} - \widehat{y}| = |\widetilde{y} - f_\lambda(\widetilde{x})|$ for regression.

$$\pi(\lambda) = \{\{i \in \mathcal{V} : \widetilde{y}_i - f_\lambda(\widetilde{x}_i) \geq 0\},$$
$$\{i \in \mathcal{V} : \widetilde{y}_i - f_\lambda(\widetilde{x}_i) < 0\}\} \stackrel{\text{def}}{=} \{\mathcal{I}_+(\lambda), \mathcal{I}_-(\lambda)\}$$

Define an invariant region of λ, which has a same partition $\pi(\lambda_0)$,

$$\mathcal{IR}(\lambda_0) = \{\lambda \in [a_{k-1}, a_k] : \pi(\lambda) = \pi(\lambda_0)\}$$

Theorem

The $\mathcal{IR}(\lambda_0)$ is a convex set and its closure is a nontrivial interval region.

Piecewise Linear

The absolute error (loss) function $L(\widetilde{y}, \widehat{y}) = |\widetilde{y} - \widehat{y}| = |\widetilde{y} - f_\lambda(\widetilde{x})|$ for regression.

$$\pi(\lambda) = \{\{i \in \mathcal{V} : \widetilde{y}_i - f_\lambda(\widetilde{x}_i) \geq 0\},$$
$$\{i \in \mathcal{V} : \widetilde{y}_i - f_\lambda(\widetilde{x}_i) < 0\}\} \stackrel{\text{def}}{=} \{\mathcal{I}_+(\lambda), \mathcal{I}_-(\lambda)\}$$

Define an invariant region of λ, which has a same partition $\pi(\lambda_0)$,

$$\mathcal{IR}(\lambda_0) = \{\lambda \in [a_{k-1}, a_k] : \pi(\lambda) = \pi(\lambda_0)\}$$

Theorem

The $\mathcal{IR}(\lambda_0)$ is a convex set and its closure is a nontrivial interval region.

Piecewise Linear

The absolute error (loss) function $L(\widetilde{y}, \widehat{y}) = |\widetilde{y} - \widehat{y}| = |\widetilde{y} - f_\lambda(\widetilde{x})|$ for regression.

$$\pi(\lambda) = \{\{i \in \mathcal{V} : \widetilde{y}_i - f_\lambda(\widetilde{x}_i) \geq 0\},$$
$$\{i \in \mathcal{V} : \widetilde{y}_i - f_\lambda(\widetilde{x}_i) < 0\}\} \stackrel{\text{def}}{=} \{\mathcal{I}_+(\lambda), \mathcal{I}_-(\lambda)\}$$

Define an invariant region of λ, which has a same partition $\pi(\lambda_0)$,

$$\mathcal{IR}(\lambda_0) = \{\lambda \in [a_{k-1}, a_k] : \pi(\lambda) = \pi(\lambda_0)\}$$

Theorem

The $\mathcal{IR}(\lambda_0)$ is a convex set and its closure is a nontrivial interval region.

Piecewise Linear

The absolute error (loss) function $L(\widetilde{y}, \widehat{y}) = |\widetilde{y} - \widehat{y}| = |\widetilde{y} - f_\lambda(\widetilde{x})|$ for regression.

$$\pi(\lambda) = \{\{i \in \mathcal{V} : \widetilde{y}_i - f_\lambda(\widetilde{x}_i) \geq 0\},$$
$$\{i \in \mathcal{V} : \widetilde{y}_i - f_\lambda(\widetilde{x}_i) < 0\}\} \stackrel{\text{def}}{=} \{\mathcal{I}_+(\lambda), \mathcal{I}_-(\lambda)\}$$

Define an invariant region of λ, which has a same partition $\pi(\lambda_0)$,

$$\mathcal{IR}(\lambda_0) = \{\lambda \in [a_{k-1}, a_k] : \pi(\lambda) = \pi(\lambda_0)\}$$

Theorem

The $\mathcal{IR}(\lambda_0)$ is a convex set and its closure is a nontrivial interval region.

Piecewise Linear

$$E(\hat{\alpha}(\lambda), \mathcal{V}, L) = \frac{1}{\ell} \sum_{i=1}^{\ell} |\widetilde{y}_i - f_\lambda(\widetilde{x}_i)| \tag{2}$$

$$= \frac{1}{\ell}\Big(\sum_{i \in \mathcal{I}_+(\lambda_0)} (\widetilde{y}_i - f_\lambda(\widetilde{x}_i)) - \sum_{i \in \mathcal{I}_-(\lambda_0)} (\widetilde{y}_i - f_\lambda(\widetilde{x}_i)) \Big)$$

$$= \frac{1}{\ell}\Big(\sum_{i \in \mathcal{I}_-(\lambda_0)} (\widetilde{y}_i - f_{\lambda_0}(\widetilde{x}_i) + \gamma_i^{[k]} \lambda_0) - \sum_{i \in \mathcal{I}_+(\lambda_0)} (\widetilde{y}_i - f_{\lambda_0}(\widetilde{x}_i) + \gamma_i^{[k]} \lambda_0) \Big)$$

$$+ \frac{1}{\ell} \left(\sum_{i \in \mathcal{I}_-(\lambda_0)} \gamma_i^{[k]} - \sum_{i \in \mathcal{I}_+(\lambda_0)} \gamma_i^{[k]} \right) \cdot \lambda$$

The mean absolute error (MAE) is *piecewise linear* w.r.t. λ, in the whole interval $[a, b]$

Piecewise Linear

$$E(\hat{\alpha}(\lambda), \mathcal{V}, L) = \frac{1}{\ell} \sum_{i=1}^{\ell} |\widetilde{y}_i - f_\lambda(\widetilde{x}_i)| \tag{2}$$

$$= \frac{1}{\ell}\Big(\sum_{i \in \mathcal{I}_+(\lambda_0)} (\widetilde{y}_i - f_\lambda(\widetilde{x}_i)) - \sum_{i \in \mathcal{I}_-(\lambda_0)} (\widetilde{y}_i - f_\lambda(\widetilde{x}_i)) \Big)$$

$$= \frac{1}{\ell}\Big(\sum_{i \in \mathcal{I}_-(\lambda_0)} (\widetilde{y}_i - f_{\lambda_0}(\widetilde{x}_i) + \gamma_i^{[k]} \lambda_0) - \sum_{i \in \mathcal{I}_+(\lambda_0)} (\widetilde{y}_i - f_{\lambda_0}(\widetilde{x}_i) + \gamma_i^{[k]} \lambda_0) \Big)$$

$$+ \frac{1}{\ell} \left(\sum_{i \in \mathcal{I}_-(\lambda_0)} \gamma_i^{[k]} - \sum_{i \in \mathcal{I}_+(\lambda_0)} \gamma_i^{[k]} \right) \cdot \lambda$$

The mean absolute error (MAE) is *piecewise linear* w.r.t. λ, in the whole interval $[a, b]$

Piecewise Constant

The standard loss function $L(\widetilde{y}, \widehat{y}) = \frac{1}{2}|\widetilde{y} - \widehat{y}|$ for binary classification, where $\widehat{y} = \text{sign}(f_\lambda(\widetilde{x}))$.

$$\begin{aligned} \widetilde{\pi}(\lambda) &= \{\{i \in \mathcal{V} : f_\lambda(\widetilde{x}_i) \geq 0\}, \{i \in \mathcal{V} : f_\lambda(\widetilde{x}_i) < 0\}\} \\ &\stackrel{\text{def}}{=} \left\{\widetilde{\mathcal{I}}_+(\lambda), \widetilde{\mathcal{I}}_-(\lambda)\right\} \end{aligned}$$

Define an invariant region of λ, which has a same partition $\pi(\lambda_0)$,

$$\widetilde{\mathcal{IR}}(\lambda_0) = \{\lambda \in [a_{k-1}, a_k] : \widetilde{\pi}(\lambda) = \widetilde{\pi}(\lambda_0)\}$$

Theorem

The $\widetilde{\mathcal{IR}}(\lambda_0)$ is a convex set and its closure is a nontrivial interval region.

Piecewise Constant

The standard loss function $L(\widetilde{y}, \widehat{y}) = \frac{1}{2}|\widetilde{y} - \widehat{y}|$ for binary classification, where $\widehat{y} = \text{sign}(f_\lambda(\widetilde{x}))$.

$$\begin{aligned} \widetilde{\pi}(\lambda) &= \{\{i \in \mathcal{V} : f_\lambda(\widetilde{x}_i) \geq 0\}, \{i \in \mathcal{V} : f_\lambda(\widetilde{x}_i) < 0\}\} \\ &\stackrel{\text{def}}{=} \left\{\widetilde{\mathcal{I}}_+(\lambda), \widetilde{\mathcal{I}}_-(\lambda)\right\} \end{aligned}$$

Define an invariant region of λ, which has a same partition $\pi(\lambda_0)$,

$$\widetilde{\mathcal{IR}}(\lambda_0) = \{\lambda \in [a_{k-1}, a_k] : \widetilde{\pi}(\lambda) = \widetilde{\pi}(\lambda_0)\}$$

Theorem

The $\widetilde{\mathcal{IR}}(\lambda_0)$ is a convex set and its closure is a nontrivial interval region.

Piecewise Constant

The standard loss function $L(\widetilde{y}, \widehat{y}) = \frac{1}{2}|\widetilde{y} - \widehat{y}|$ for binary classification, where $\widehat{y} = \text{sign}(f_\lambda(\widetilde{x}))$.

$$\begin{aligned} \widetilde{\pi}(\lambda) &= \{\{i \in \mathcal{V} : f_\lambda(\widetilde{x}_i) \geq 0\}, \{i \in \mathcal{V} : f_\lambda(\widetilde{x}_i) < 0\}\} \\ &\stackrel{\text{def}}{=} \left\{\widetilde{\mathcal{I}}_+(\lambda), \widetilde{\mathcal{I}}_-(\lambda)\right\} \end{aligned}$$

Define an invariant region of λ, which has a same partition $\pi(\lambda_0)$,

$$\widetilde{\mathcal{IR}}(\lambda_0) = \{\lambda \in [a_{k-1}, a_k] : \widetilde{\pi}(\lambda) = \widetilde{\pi}(\lambda_0)\}$$

Theorem

The $\widetilde{\mathcal{IR}}(\lambda_0)$ is a convex set and its closure is a nontrivial interval region.

Piecewise Constant

The standard loss function $L(\widetilde{y}, \widehat{y}) = \frac{1}{2}|\widetilde{y} - \widehat{y}|$ for binary classification, where $\widehat{y} = \text{sign}(f_\lambda(\widetilde{x}))$.

$$\begin{aligned} \widetilde{\pi}(\lambda) &= \{\{i \in \mathcal{V} : f_\lambda(\widetilde{x}_i) \geq 0\}, \{i \in \mathcal{V} : f_\lambda(\widetilde{x}_i) < 0\}\} \\ &\stackrel{\text{def}}{=} \left\{\widetilde{\mathcal{I}}_+(\lambda), \widetilde{\mathcal{I}}_-(\lambda)\right\} \end{aligned}$$

Define an invariant region of λ, which has a same partition $\pi(\lambda_0)$,

$$\widetilde{\mathcal{IR}}(\lambda_0) = \{\lambda \in [a_{k-1}, a_k] : \widetilde{\pi}(\lambda) = \widetilde{\pi}(\lambda_0)\}$$

Theorem

The $\widetilde{\mathcal{IR}}(\lambda_0)$ is a convex set and its closure is a nontrivial interval region.

Piecewise Constant

$$E(\hat{\alpha}(\lambda), \mathcal{V}, L) = \frac{1}{2\ell} \sum_{i=1}^{\ell} |\widetilde{y}_i - \text{sign}\left(f_\lambda(\widetilde{x}_i)\right)| \tag{3}$$

$$= \frac{1}{2\ell} \left(\sum_{i \in \mathcal{I}_+(\lambda_0)} |\widetilde{y}_i - 1| + \sum_{i \in \mathcal{I}_-(\lambda_0)} |\widetilde{y}_i + 1| \right)$$

The error rate (ER) is *piecewise constant* w.r.t. λ, in the whole interval $[a, b]$

1. The weighted error rate
2. The precision,
3. The recall
4. The F-measure

Piecewise Constant

$$E(\hat{\alpha}(\lambda), \mathcal{V}, L) = \frac{1}{2\ell} \sum_{i=1}^{\ell} |\widetilde{y}_i - \text{sign}\left(f_\lambda(\widetilde{x}_i)\right)| \tag{3}$$

$$= \frac{1}{2\ell} \left(\sum_{i \in \mathcal{I}_+(\lambda_0)} |\widetilde{y}_i - 1| + \sum_{i \in \mathcal{I}_-(\lambda_0)} |\widetilde{y}_i + 1| \right)$$

The error rate (ER) is *piecewise constant* w.r.t. λ, in the whole interval $[a, b]$

1. The weighted error rate
2. The precision,
3. The recall
4. The F-measure

Piecewise Constant

$$E(\hat{\alpha}(\lambda), \mathcal{V}, L) = \frac{1}{2\ell} \sum_{i=1}^{\ell} |\widetilde{y}_i - \text{sign}\left(f_\lambda(\widetilde{x}_i)\right)| \tag{3}$$

$$= \frac{1}{2\ell} \left(\sum_{i \in \mathcal{I}_+(\lambda_0)} |\widetilde{y}_i - 1| + \sum_{i \in \mathcal{I}_-(\lambda_0)} |\widetilde{y}_i + 1| \right)$$

The error rate (ER) is *piecewise constant* w.r.t. λ, in the whole interval $[a, b]$

1. The weighted error rate
2. The precision,
3. The recall
4. The F-measure

Computing Generalized Error Path

To fit various error paths, we define a 2-tuple (I, P)

1. For piecewise quadratic error path:
   $I = [a_{k-1}, a_k]$, $P.c_2 = \frac{1}{\ell}\sum_{i=1}^{\ell}(\gamma_i^{[k]})^2$,
   $P.c_1 = -\frac{1}{\ell}\sum_{i=1}^{\ell} 2 \times \gamma_i^{[k]}(\widetilde{y}_i - f_{a_{k-1}}(\widetilde{x}_i) + \gamma_i^{[k]} a_{k-1})$,
   $P.c_0 = \frac{1}{\ell}\sum_{i=1}^{\ell}\left(\widetilde{y}_i - f_{a_{k-1}}(\widetilde{x}_i) + \gamma_i^{[k]} a_{k-1}\right)^2$.
2. For piecewise linear error path:
   $I = \mathcal{IR}(\lambda_0)$,
   $P.c_1 = \frac{1}{\ell}\left(\sum_{i\in\mathcal{I}_-(\lambda_0)}\gamma_i^{[k]} - \sum_{i\in\mathcal{I}_+(\lambda_0)}\gamma_i^{[k]}\right)$,
   $P.c_0 = \frac{1}{\ell}(\sum_{i\in\mathcal{I}_-(\lambda_0)}(\widetilde{y}_i - f_{\lambda_0}(\widetilde{x}_i) + \gamma_k\lambda_0) - \sum_{i\in\mathcal{I}_+(\lambda_0)}(\widetilde{y}_i - f_{\lambda_0}(\widetilde{x}_i) + \gamma_k\lambda_0))$.
3. For piecewise constant error path:
   $I = \widetilde{\mathcal{IR}}(\lambda_0)$, $P.c_0 = E(\hat{\alpha}(\lambda_0), \mathcal{V}, L)$.

GEP (Generalized error path algorithm)

Input: A solution path w.r.t. λ in an interval $[a, b] = \bigcup_{k=1}^{m} [a_{k-1}, a_k]$, a validation set $\mathcal{V}$.

Output: An error path $\{(I_1, P_1), (I_2, P_2), (I_3, P_3), \cdots\}$.

Initialize $\lambda = a$, $k = 0$, $j = 1$.
while $\lambda < b$ **do**
  Update $k = k + 1$, read the k-th sub-interval $[a_{k-1}, a_k]$ from the solution path.
  while $\lambda < a_k$ **do**
    Compute (I_j, P_j) for the leftmost in $[\lambda, a_k]$.
    Update $\lambda = R(I_j)$, $j = j + 1$.
  end while
end while

CV-GEP (Cross validation with GEP)

Input: K error paths $\{(I_1^k, P_1^k), (I_2^k, P_2^k), (I_3^k, P_3^k), \cdots\}_{k=1}^K$ in the interval $[a, b]$.

Output: An error path $\{(I_1, P_1), (I_2, P_2), (I_3, P_3), \cdots\}$.

1: Initialize $\lambda = a$, $i^1 = 1, \cdots, i^K = 1$, $j = 1$.
2: **while** $\lambda < b$ **do**
3: Compute (I_j, P_j) from $(I_{i^1}^1, P_{i^1}^1), \cdots, (I_{i^K}^K, P_{i^K}^K)$.
4: Update $i^1, \cdots, i^K$. (If $R(I_j) = R(I_{i^k}^k)$, update $i^k = i^k + 1$.)
5: Update $\lambda = R(I_j)$, and $j = j + 1$.
6: **end while**

Datasets

Table: Summary of datasets. (BC=binary classification, R=regression)

Dataset	Number of samples	Dimensionality	Task
Ionosphere	354	34	BC
Diabetes	768	8	BC
Hill-Valley	606	100	BC
Breast Cancer	683	10	BC
Spine	350	5	BC
Friedman	1,500	10	R
Housing	506	13	R
Forest Fires	517	12	R
Auto MPG	392	7	R
Triazines	186	60	R

Table: The results of CV error obtained from GS, MS, RS, and our GEP.

Dataset	ER				Dataset	MAE				MSE			
	GS	MS	RS	GEP		GS	MS	RS	GEP	GS	MS	RS	GEP
Ionosphere	0.062	0.068	0.068	**0.060**	Friedman	2.08	2.1	2.07	**1.98**	6.81	6.82	6.82	**6.60**
Diabetes	0.655	0.655	0.565	**0.345**	Housing	2.12	2.13	2.13	**2.09**	8.11	8.14	8.14	**8.05**
Hill-Valley	0.465	0.470	0.468	**0.460**	Forest Fires	17.9	18.3	18.2	**17.2**	3251	3258	3263	**3238**
Breast Cancer	0.528	0.528	0.474	**0.342**	Auto MPG	2.39	2.41	2.39	**2.36**	9.62	9.63	9.62	**9.58**
Spine	0.060	0.058	0.059	**0.055**	Triazines	2.39	2.40	2.40	**2.37**	9.66	9.67	9.66	**9.63**

Table: The average errors on the test data, over 10 trails, obtained from GS, MS, RS, and our GEP.

Dataset	ER				Dataset	MAE				MSE			
	GS	MS	RS	GEP		GS	MS	RS	GEP	GS	MS	RS	GEP
Ionosphere	0.066	0.066	0.065	**0.063**	Friedman	2.12	2.16	2.14	**2.08**	6.84	6.83	6.84	**6.75**
Diabetes	0.356	0.356	0.358	**0.347**	Housing	2.05	2.06	2.06	**2.05**	8.56	8.63	8.66	**8.44**
Hill-Valley	0.514	0.519	0.514	**0.512**	Forest Fires	18.4	19.1	18.9	**17.9**	5468	5475	5471	**5446**
Breast Cancer	0.528	0.528	0.474	**0.342**	Auto MPG	2.74	2.75	2.76	**2.62**	14.6	14.6	14.7	**14.0**
Spine	0.060	0.061	0.061	**0.058**	Triazines	2.76	2.77	2.77	**2.54**	14.7	14.8	14.6	**14.2**

Table: The results of CV error obtained from GS, MS, RS, and our GEP.

Dataset	ER				Dataset	MAE				MSE			
	GS	MS	RS	GEP		GS	MS	RS	GEP	GS	MS	RS	GEP
Ionosphere	0.062	0.068	0.068	**0.060**	Friedman	2.08	2.1	2.07	**1.98**	6.81	6.82	6.82	**6.60**
Diabetes	0.655	0.655	0.565	**0.345**	Housing	2.12	2.13	2.13	**2.09**	8.11	8.14	8.14	**8.05**
Hill-Valley	0.465	0.470	0.468	**0.460**	Forest Fires	17.9	18.3	18.2	**17.2**	3251	3258	3263	**3238**
Breast Cancer	0.528	0.528	0.474	**0.342**	Auto MPG	2.39	2.41	2.39	**2.36**	9.62	9.63	9.62	**9.58**
Spine	0.060	0.058	0.059	**0.055**	Triazines	2.39	2.40	2.40	**2.37**	9.66	9.67	9.66	**9.63**

Table: The average errors on the test data, over 10 trails, obtained from GS, MS, RS, and our GEP.

Dataset	ER				Dataset	MAE				MSE			
	GS	MS	RS	GEP		GS	MS	RS	GEP	GS	MS	RS	GEP
Ionosphere	0.066	0.066	0.065	**0.063**	Friedman	2.12	2.16	2.14	**2.08**	6.84	6.83	6.84	**6.75**
Diabetes	0.356	0.356	0.358	**0.347**	Housing	2.05	2.06	2.06	**2.05**	8.56	8.63	8.66	**8.44**
Hill-Valley	0.514	0.519	0.514	**0.512**	Forest Fires	18.4	19.1	18.9	**17.9**	5468	5475	5471	**5446**
Breast Cancer	0.528	0.528	0.474	**0.342**	Auto MPG	2.74	2.75	2.76	**2.62**	14.6	14.6	14.7	**14.0**
Spine	0.060	0.061	0.061	**0.058**	Triazines	2.76	2.77	2.77	**2.54**	14.7	14.8	14.6	**14.2**

Conclusion

1. Propose a new generalized error path algorithm (GEP)
   - incorporating more general error (or loss) functions
   - working with more general solution path algorithms.
   - obtaining minimum CV errors directly for a large variety of error (or loss) functions and solution path algorithms
2. Other learning problems, such as, ordinal regression, ranking, multiple classification, ...
3. Other loss functions, such as AUC, ...

Thank You!