On Discriminative Learning of Prediction Uncertainty

Vojtěch Franc, Daniel Průša

Department of Cybernetics
Faculty of Electrical Engineering
Czech Technical University in Prague

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION,
YOUTH AND SPORTS

CTU
CZECH TECHNICAL UNIVERSITY IN PRAGUE

RESEARCH CENTER FOR INFORMATICS
rci.cvut.cz

On Discriminative Learning of Prediction Uncertainty

Selective classifier: $(h,c)(x) = \begin{cases} h(x) & \text{with probability} \quad c(x) \\ \text{reject} & \text{with probability} \quad 1-c(x) \end{cases}$

where $h\colon \mathcal{X} \to \mathcal{Y}$ is a classifier and $c\colon \mathcal{X} \to [0,1]$ is a selection function

Selective classifier: $(h,c)(x) = \begin{cases} h(x) & \text{with probability} \quad c(x) \\ \text{reject} & \text{with probability} \quad 1 - c(x) \end{cases}$

where $h\colon \mathcal{X} \to \mathcal{Y}$ is a classifier and $c\colon \mathcal{X} \to [0,1]$ is a selection function

Example: Linear SVM

$$h(x) = \operatorname{sign}(\langle \boldsymbol{\phi}(x), \boldsymbol{w} \rangle + b)$$

$$c(x) = [\![|\langle \boldsymbol{\phi}(x), \boldsymbol{w} \rangle + b| \geq \theta]\!]$$

Selective classifier: $(h,c)(x) = \begin{cases} h(x) & \text{with probability} \quad c(x) \\ \text{reject} & \text{with probability} \quad 1-c(x) \end{cases}$

where $h\colon \mathcal{X} \to \mathcal{Y}$ is a classifier and $c\colon \mathcal{X} \to [0,1]$ is a selection function

Example: Linear SVM

$$h(x) = \operatorname{sign}(\langle \boldsymbol{\phi}(x), \boldsymbol{w}\rangle + b)$$

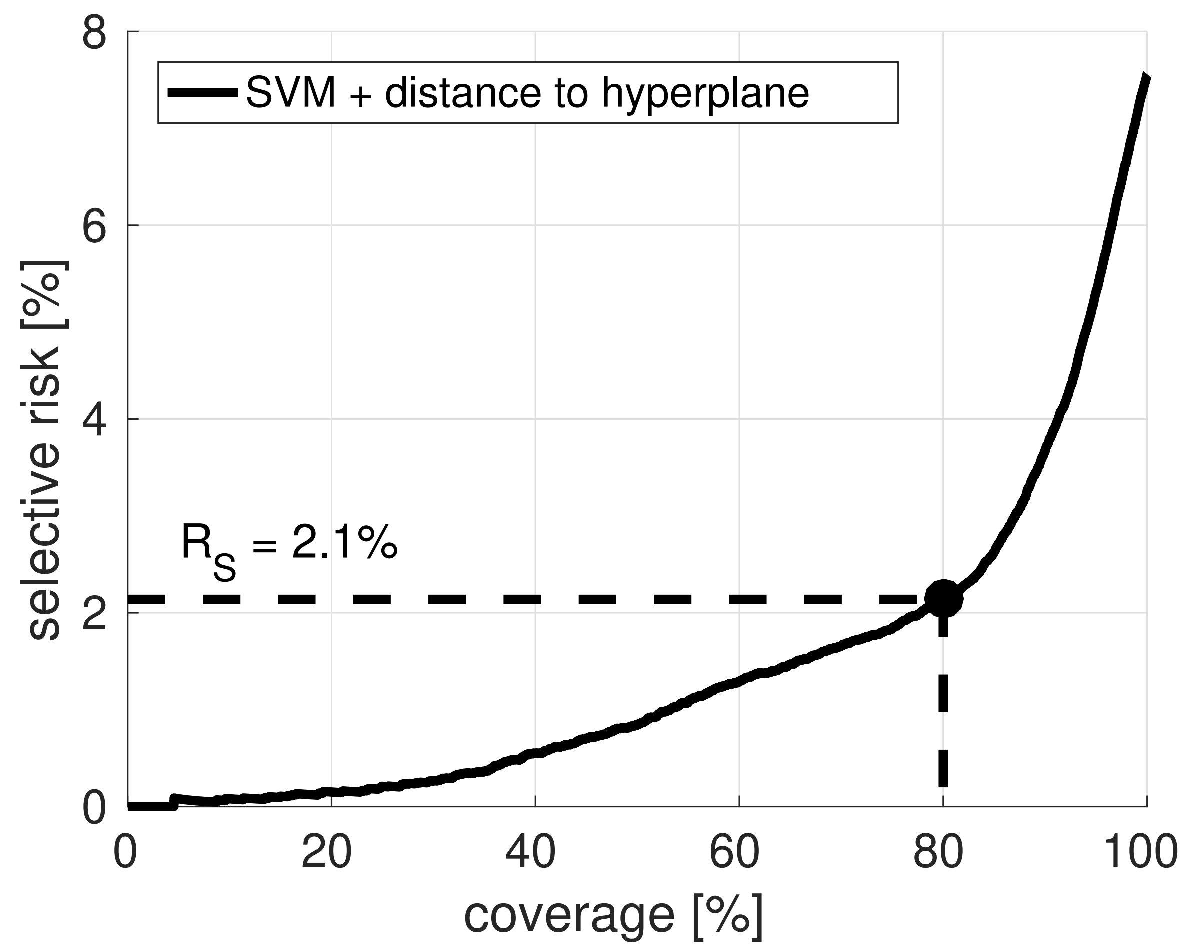
$$c(x) = [\![|\langle \boldsymbol{\phi}(x), \boldsymbol{w}\rangle + b| \geq \theta]\!]$$

Coverage:

$$\phi(c) = \mathbb{E}_{x\sim p}\big[c(x)\big]$$

Selective risk:

$$R_S(h,c) = \frac{\mathbb{E}_{(x,y)\sim p}\Big[\ell(y,h(x))\, c(x)\Big]}{\phi(x)}$$



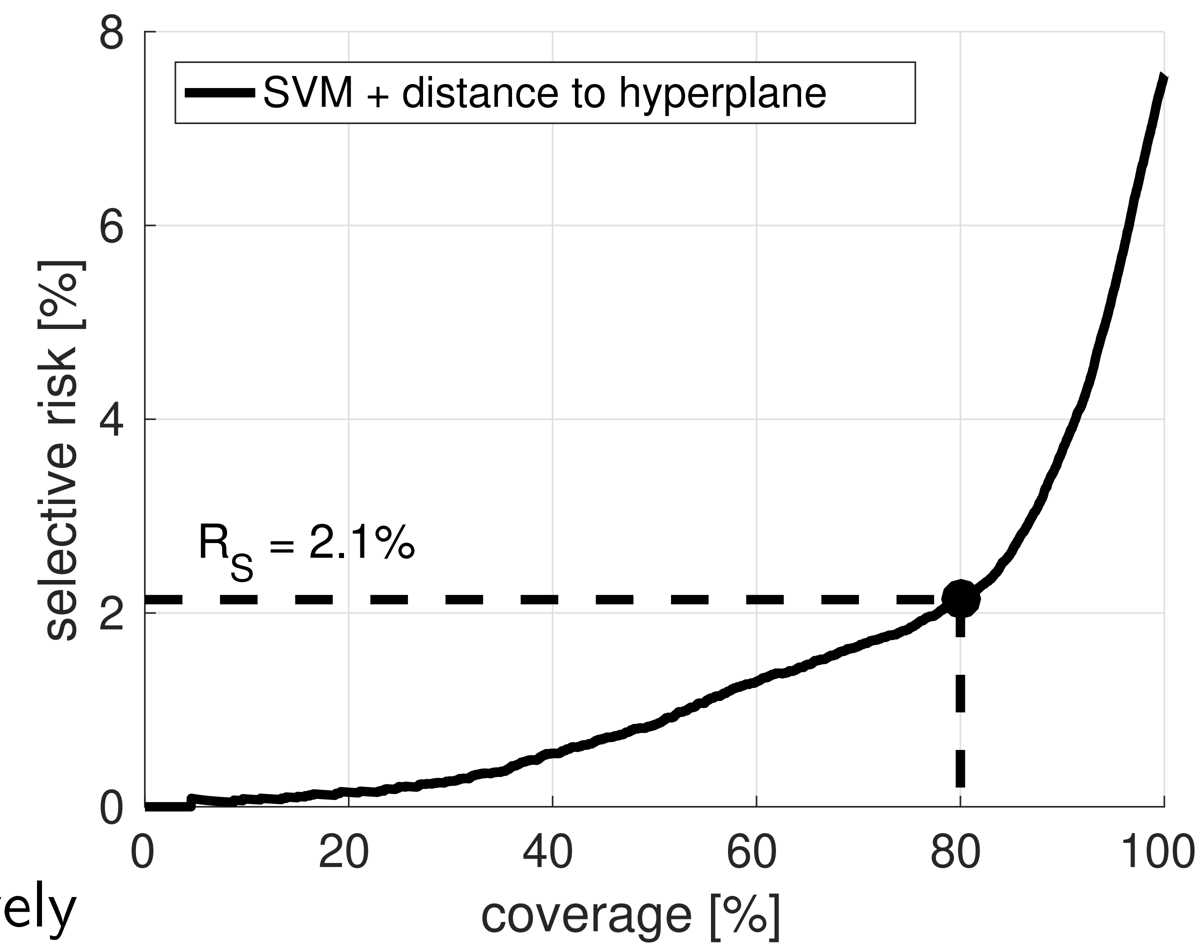
On Discriminative Learning of Prediction Uncertainty

Selective classifier: $(h,c)(x) = \begin{cases} h(x) & \text{with probability} \quad c(x) \\ \text{reject} & \text{with probability} \quad 1 - c(x) \end{cases}$

where $h\colon \mathcal{X} \to \mathcal{Y}$ is a classifier and $c\colon \mathcal{X} \to [0,1]$ is a selection function

Example: Linear SVM

$$h(x) = \text{sign}(\langle \boldsymbol{\phi}(x), \boldsymbol{w} \rangle + b)$$
$$c(x) = [\![|\langle \boldsymbol{\phi}(x), \boldsymbol{w} \rangle + b| \geq \theta]\!]$$



In our paper we show:

1) What is the optimal $c(x)$

2) How to learn $c(x)$ discriminatively

On Discriminative Learning of Prediction Uncertainty

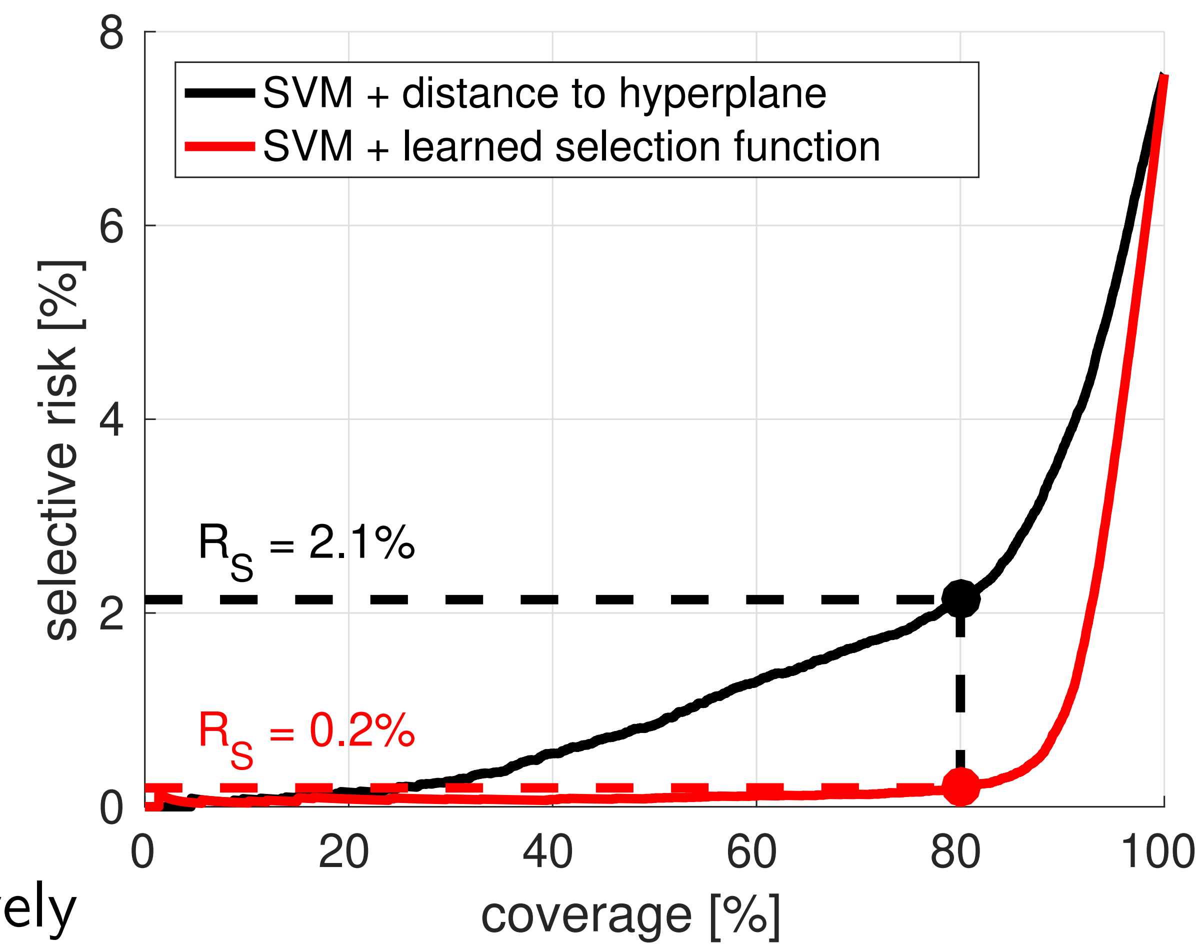
Selective classifier: $(h, c)(x) = \begin{cases} h(x) & \text{with probability} \quad c(x) \\ \text{reject} & \text{with probability} \quad 1 - c(x) \end{cases}$

where $h\colon \mathcal{X} \to \mathcal{Y}$ is a classifier and $c\colon \mathcal{X} \to [0, 1]$ is a selection function

Example: Linear SVM

$$h(x) = \text{sign}(\langle \boldsymbol{\phi}(x), \boldsymbol{w} \rangle + b)$$

$$c(x) = [\![|\langle \boldsymbol{\phi}(x), \boldsymbol{w} \rangle + b| \geq \theta]\!]$$



In our paper we show:

1) What is the optimal $c(x)$

2) How to learn $c(x)$ discriminatively