Table 1: Comparison of region-specific and locally linear models.

	Ours	LSL-SP	CSTC	LDKL	FaLK-SVM	LLSVM
Region Optimization	√	√	√	√		
Initialization-independent	√					√
Generalization Bound	√	√			√	
Region Specifiers	(Sec. 2.2)	Linear	Linear	Linear	Non-Regional	Non-Regional

1.2 Notations

Scalars and vectors are denoted by lower-case x. Matrices are denoted by upper-case X. An n-th training sample and label are denoted by $x_n \in \mathbb{R}^D$ and y_n, respectively.

2 Partition-wise Linear Models

This section explains partition-wise linear models under the assumption that effective partitioning is already fixed. We discuss how to optimize partitions and region-specific linear models in Section 3.

2.1 Framework

Figure 1 illustrates the concept of partition-wise linear models. Suppose we have P partitions (red dashed lines) which essentially specify 2^P regions. Partition-wise linear models are defined as follows. First, we assign a linear weight vector a_p to the p-th partition. This partition has an *activeness function*, f_p, which indicates whether the attached weight vector a_p is applied to individual data points or not. For example, in Figure 1, we set the weight vector a_1 to be applied to the right-hand side of partition p_1. In this case, the corresponding activeness function is defined as $f_1(x) = 1$ when x is in the right-hand side of p_1. Second, region-specific predictors (squared regions surrounded by partitions in Figure 1) are defined by a linear combination of active partition-wise weight vectors that are also linear models.

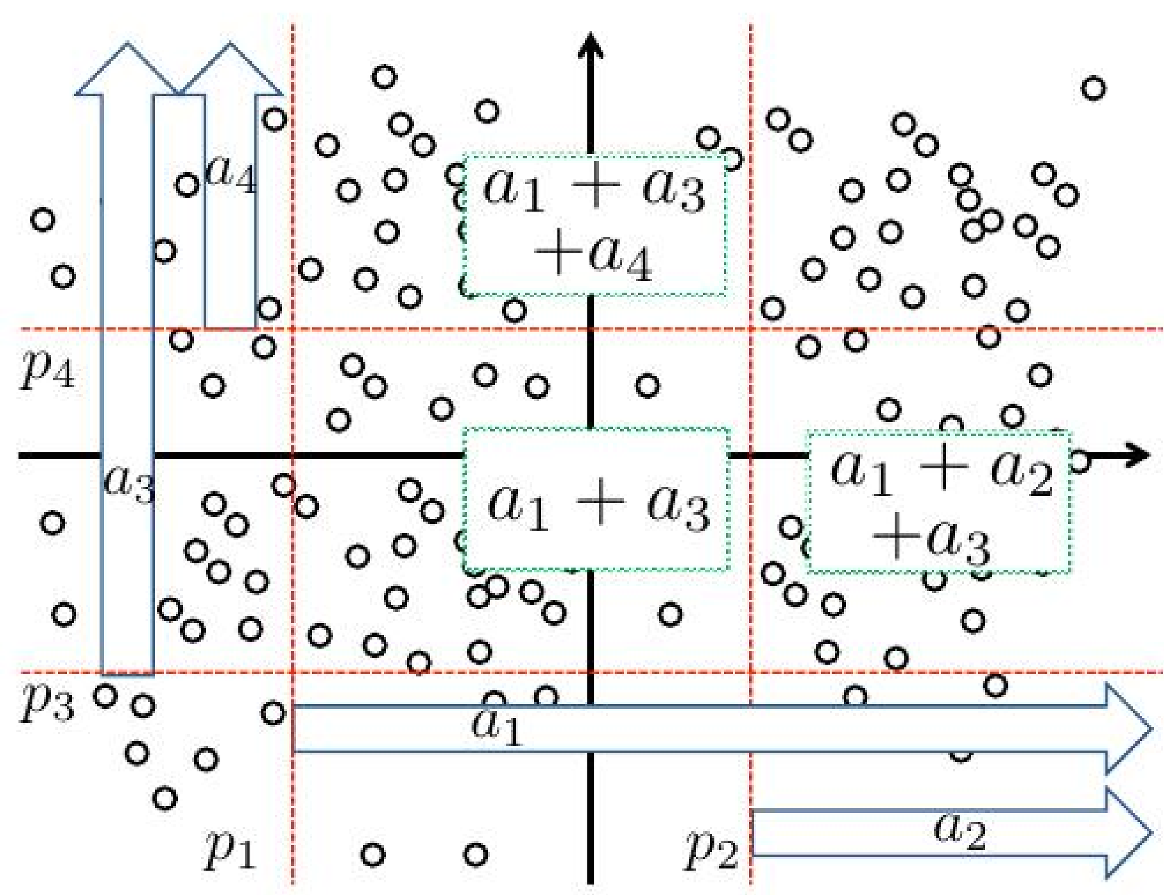

Figure 1: Concept of Partition-wise Linear Models

Let us formally define the partition-wise linear models. We have a set of given activeness functions, $f_1, \dots, f_P$, which is denoted in a vector form as $f(\cdot) = (f_1(\cdot), \dots, f_P(\cdot))^T$. The p-th element $f_p(x) \in \{0, 1\}$ indicates whether the attached weight vector a_p is applied to x or not. The activeness function $f(\cdot)$ can represent at most 2^P regions, and $f(x)$ specifies to which region x belongs. A linear model of an individual region is then represented as $\sum_{p=1}^{P} f_p(\cdot) a_p$. It is worth noting that partition-wise linear models use P linear weight vectors to represent 2^P regions and restrict the number of parameters.

The overall predictor $g(\cdot)$ can be denoted as follows:

$$g(x) = \sum_p f_p(x) \sum_d a_{dp} x_d. \tag{1}$$

Let us define A as $A = (a_1, \dots, a_P)$. The partition-wise linear model (1) simply acts as a linear model w.r.t. A while it captures the non-linear nature of data (individual regions use different linear models). Such non-linearity originates from the activeness functions f_ps, which are fundamentally important components in our models.

By introducing a convex loss function $\ell(\cdot, \cdot)$ (e.g., squared loss for regression, squared hinge or logistic loss for classification), we can represent an objective function of the partition-wise linear models as a convex loss minimization problem as follows:

$$\min_A \sum_n \ell(y_n, g(x_n)) = \min_A \sum_n \ell(y_n, \sum_p f_p(x_n) \sum_d a_{dp} x_{nd}). \tag{2}$$

Here we give a convex formulation of region-specific linear models under the assumption that a set of partitions is given. In Section 3, we propose a convex optimization algorithm for partitions and regions as a partition selection problem, using sparsity-inducing structured regularization.