**Algorithm 2** Selection Algorithm, $\mathcal{A}$

**Input** Training data with labels $(\boldsymbol{y}, X)$. A kernel functions $K$. A number $s$, the total size of the bases. (Optional)

1. Get a set of basis, $B^{big}$, each with size $\sqrt{\ell}$. Randomly sample $\sqrt{\ell}$ instances with labels, $(\boldsymbol{y}^{subset}, X^{subset})$.
2. $X^{select} \leftarrow K(X^{subset}, B^{big})$.
3. Train L1-regularized SVM on$(\boldsymbol{y}^{subset}, X^{select})$ and a sparsity pattern $\boldsymbol{w}$ for each class.
4. $B \leftarrow$ columns in $B^{big}$ where the corresponding element in $\boldsymbol{w}$ is not zero.
5. If $s$ is defined, $B \leftarrow$ the $s$ centroid generated by k-means of $B_i$ (as proposed in [12].)

**Return** $B$