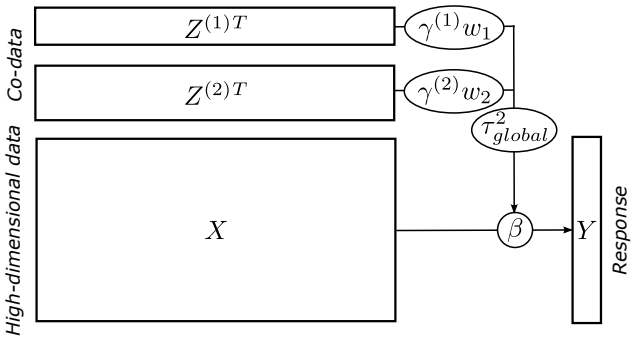
Main workflow

Cheat sheet for `R`*-package* `ecpc`*:*
Empirical bayes Co-data learnt Prediction and Covariate selection

- Install and load package:

```
install.packages("ecpc")
library("ecpc")
```

- Prepare data and co-data (see right):



- Estimate parameters:

```
fit <- ecpc(Y, X, Z=list(Z_1, Z_2))
```

- Visualise estimates:

>Regression coefficients:

```
plot(fit, show="coefficients")
```

>Prior variances:

```
plot(fit, show="priorweights")
```

- Predict for new samples X_2:

```
predictions <- predict(fit, X_2)
```

- Select variables:

>A posteriori:

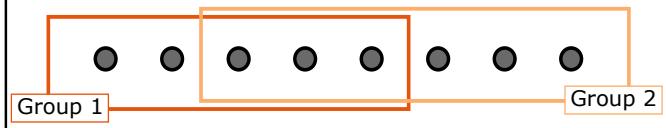
```
fit_post <- postSelect(fit, X, Y)
```

>Transform ridge to elastic net penalties with parameter alpha:

```
fit_squeezy <- squeezy(Y, X, alpha=alpha,
                       lambdas=fit$penalties)
```

Preparing co-data

In group sets



- Create group set:

>Categorical:

```
gs_1 <- createGroupset(factor)
```

>Continuous discretised in non-overlapping groups:

```
gs_2 <- createGroupset(values)
```

>Continuous discretised in overlapping groups for adaptive discretisation:

```
gs_3 <- splitMedian(values)
```

>Group set on the group level for hierarchical groups for adaptive discretisation:

```
gs_grouplvl <- obtainHierarchy(groupset_3)
```

- Choose hypershrinkage (penalty on the group level):

>Few groups:

```
hypershrinkage = "none"
```

>Many groups:

```
hypershrinkage="ridge"
```

>Select groups:

```
hypershrinkage="lasso"
```

>Groups structured in (hierarchical) groups:

```
hypershrinkage="hierLasso,ridge",
groupsets.grplvl=list(groupset_grouplvl)
```

- Estimate parameters:

```
fit_gs <- ecpc(Y, X,
               groupsets=list(gs_1, gs_2, gs_3),
               groupsets.grplvl=list(NULL, NULL,
                                     gs_grouplvl),
               hypershrinkage=c("none", "ridge",
                                "hierLasso,ridge"))
```

In co-data matrices

$$Z^{(d)} = \begin{bmatrix} 1 & 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & 1 & 1 & 1 \end{bmatrix}^T$$

- Create co-data (related) matrix:

>For group set (make dummy variables):

```
Z_1 <- createZforGroupset(gs_1)
```

>Spline matrix for continuous co-data:

```
Z_2 <- createZforSplines(values)
```

>Difference penalty matrix for splines:

```
S_1 <- createS(G=dim(Z_2)[2])
```

>Constraints for splines:

```
Con_1 <- createCon(G=dim(Z_2)[2]), shape)
```

for shape one of `"positive"`, `"monotone.i"` (increasing), `"monotone.d"` (decreasing), `"convex"`, `"concave"`, or any combination thereof by separating with a `"+"`, e.g. `"positive+convex"`

- Choose hypershrinkage (penalty on the co-data variables):

>No penalty/constraints, e.g. linear co-data model:

```
paraPen=NULL, paraCon=NULL
```

>Generalised ridge penalty, e.g. generalised additive co-data model:

```
paraPen=list(Z2=list(S1=S_1))
```

>Constraints, e.g. shape constrained additive co-data model:

```
paraCon=list(Z2=Con_1)
```

- Estimate parameters:

```
fit_Z <- ecpc(Y, X, Z=list(Z_1, Z_2),
              paraPen=list(Z2=list(S1=S_1)),
              paraCon=list(Z2=Con_1))
```