For the sake of simplicity, I will use the same dimensions of the feature set to explain stepwise regression. Note: BIC is acceptable here as well.

**AIC= 2k - 2 ln(L)**

k is the number of parameters in the model.

L is the likelihood of the model.

**BIC= ln(n)k – 2 ln(L)**

k is the number of parameters in the model.

n is the number of data points.

L is the likelihood of the model.

Forward

For a 6-feature dataset with 1 outcome and 5 potential predictors:

(1) Start with model bias and calculate the AIC value.

(2) [forward step] Try adding each of the five features to formulate five separate univariate models, and calculate AIC for each of the five models.

If AIC for any of these univariate models is less than the AIC for the model bias, then retain the univariate model with the lowest AIC and move to Step 3.

If AIC for these models is greater than AIC for the model bias, then the forward stepwise algorithm stops.

(3) [forward step] Try adding each of the remaining four features to formulate four separate bivariate models, and calculate AIC for each of the four models.

If AIC for any of these bivariate models is less than the AIC for the univariate model retained in Step 2, then retain the bivariate model with the lowest AIC and move to Step 4.

If AIC for these models is greater than AIC for the univariate model, then the forward stepwise algorithm stops and the univariate model is retained.

(4) [backward step] Try removing each of the two features from the bivariate model one-by-one so that two univariate models are compared with the bivariate model. If removing one of the features reduces AIC, then go back to Step 2. Otherwise, proceed to Step 5.

(5) [forward step] Try adding each of the remaining three features to formulate three separate trivariate models, and calculate AIC for each of the three models.

If AIC for any of these trivariate models is less than AIC for the bivariate model retained in Step 3, then retain the trivariate model with the lowest AIC and move to Step 6.

If AIC for these models is greater than AIC for the bivariate model, then the forward stepwise algorithm stops and the bivariate model is retained.

(6) [backward step] Try removing each of the three features from the trivariate model one-by-one so that three bivariate models are compared with the trivariate model. If removing one of the features reduces AIC, then go back to Step 3. Otherwise, proceed to Step 7.

(7) [continue this forward-backward process until all five variables are added or the attempt to add or remove a variable does not decrease AIC over the value for the current model]

Backward

1. Start with all 10 predictors and calculate the AIC value.
2. [backward step] Try removing each of the 10 features to formulate ten separate models and calculate the AIC for each of the ten models.

If AIC for any of these ten models is less than the AIC for the 10-predictor model, then retain the 9- predictor model and move to Step 3.

1. [backward step] Try removing each of the 9 predictors one a time to formulate nine separate models and compute the corresponding AICs. Choose the lowest AIC 8- predictor model.
2. [forward step] Try adding each of the two features model one-by-one so that two 9-parameter models are compared with the 8-predictor model. If adding one of the features reduces AIC, then go back to Step 2. Otherwise, proceed to Step 5.
3. [backward step] Try removing each of the 8 predictors one a time to formulate eight separate models and compute the corresponding AICs. Choose the lowest AIC 7- predictor model.
4. [forward step] Try adding each of the three features model one-by-one so that three 8- predictor models are compared with the 7- predictor model. If adding one of the features reduces AIC, then go back to Step 4. Otherwise, proceed to Step 7.
5. Continue this step-by-step backward-forward process till you reach a 5-predictor model.