Assessing the importance and effect of predictor variables

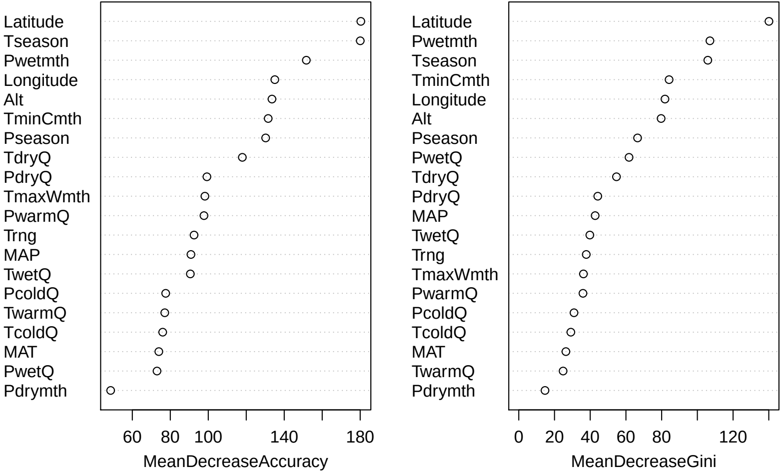
Machine Learning for Ecology workshop

SEEC - Statistics in Ecology, Environment and Conservation

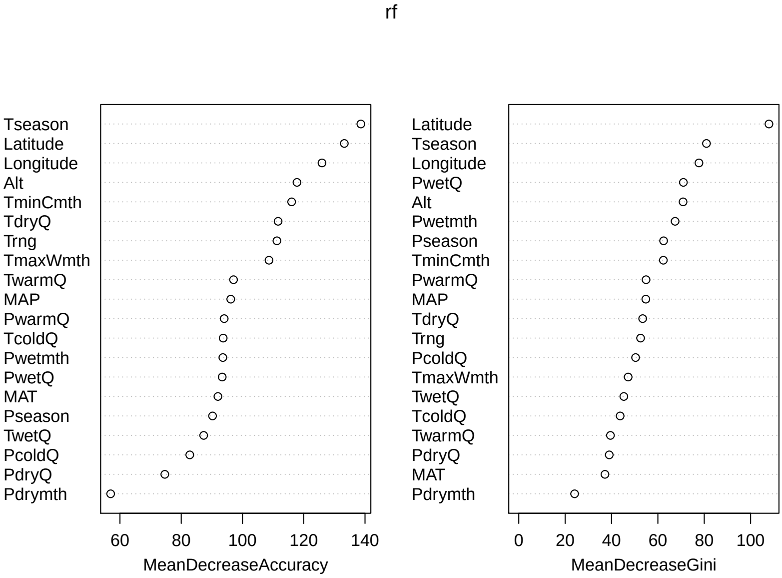
Variable Importance

- No inference with trees – no significance testing
- Variable "importance": amount by which the splitting criterion improved
- Only a *relative* measure, and no *how* information

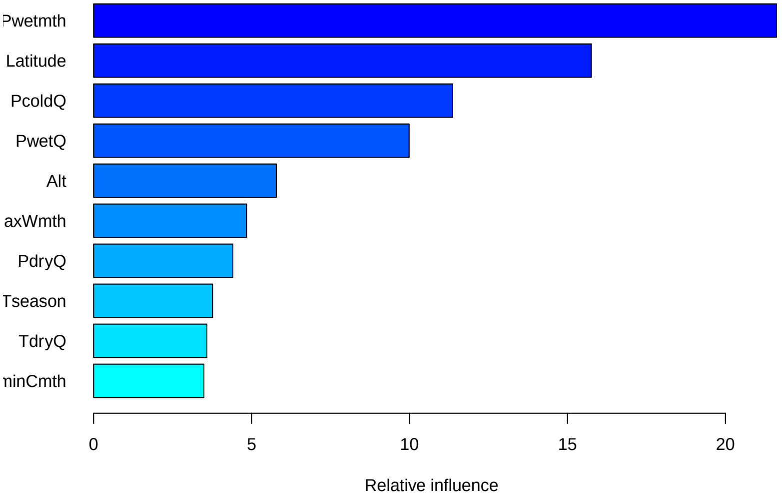
Variable importance (bagging)



Variable importance (random forest)



Variable importance (boosting)

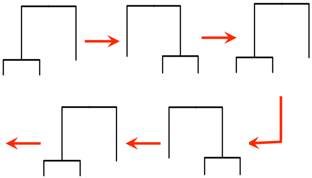


Constructing Partial Dependence Plots

Visually shows the effect of X_i on predictions *after accounting for other predictors*

	y	x1	x2	x3
1	35.70	2.17	1.77	5.78
2	52.28	2.42	5.63	6.46
3	38.18	0.78	2.74	4.36
4	35.99	0.09	3.04	3.45
5	21.19	2.21	0.50	3.40
6	54.38	-2.64	3.63	6.81
7	23.59	2.26	0.23	4.52
8	32.27	0.45	1.34	5.62
9	47.84	0.43	4.24	5.61
10	38.87	-0.84	2.60	4.84

Data



Predictive Model

Construct a partial dependence plot for X_3

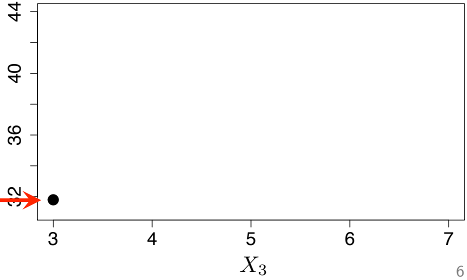
Constructing Partial Dependence Plots

Visually shows the effect of X_i on predictions *after accounting for other predictors*

	y	x1	x2	x3	yhat
1	35.70	2.17	1.77	3	25.87
2	52.28	2.42	5.63	3	42.86
3	38.18	0.78	2.74	3	32.48
4	35.99	0.09	3.04	3	34.93
5	21.19	2.21	0.50	3	20.10
6	54.38	-2.64	3.63	3	41.99
7	23.59	2.26	0.23	3	18.80
8	32.27	0.45	1.34	3	26.70
9	47.84	0.43	4.24	3	39.79
10	38.87	-0.84	2.60	3	34.44

$\hat{y}(X_3 = 3, X_{-3} = x_{-3,j})$

$\hat{y}(X_3 = 3) =$ 31.80



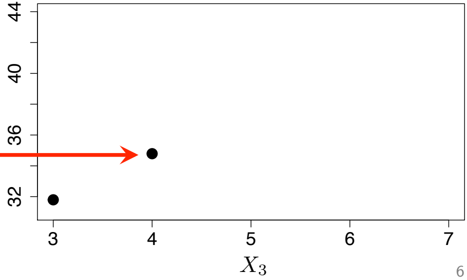
Constructing Partial Dependence Plots

Visually shows the effect of X_i on predictions *after accounting for other predictors*

	y	x1	x2	x3	yhat
1	35.70	2.17	1.77	4	28.86
2	52.28	2.42	5.63	4	45.85
3	38.18	0.78	2.74	4	35.47
4	35.99	0.09	3.04	4	37.93
5	21.19	2.21	0.50	4	23.09
6	54.38	-2.64	3.63	4	44.98
7	23.59	2.26	0.23	4	21.79
8	32.27	0.45	1.34	4	29.69
9	47.84	0.43	4.24	4	42.78
10	38.87	-0.84	2.60	4	37.43

$\hat{y}(X_3 = 4, X_{-3} = x_{-3,j})$

$\hat{y}(X_3 = 4) = 34.79$



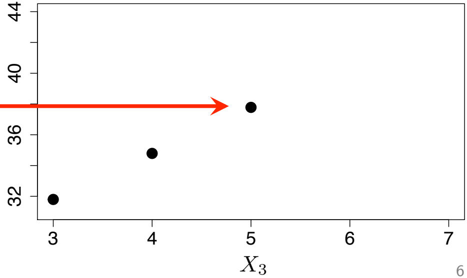
Constructing Partial Dependence Plots

Visually shows the effect of X_i on predictions *after accounting for other predictors*

	y	x1	x2	x3	yhat
1	35.70	2.17	1.77	5	31.85
2	52.28	2.42	5.63	5	48.84
3	38.18	0.78	2.74	5	38.47
4	35.99	0.09	3.04	5	40.92
5	21.19	2.21	0.50	5	26.08
6	54.38	-2.64	3.63	5	47.98
7	23.59	2.26	0.23	5	24.78
8	32.27	0.45	1.34	5	32.69
9	47.84	0.43	4.24	5	45.77
10	38.87	-0.84	2.60	5	40.42

$\hat{y}(X_3 = 5, X_{-3} = x_{-3,j})$

$\hat{y}(X_3 = 5) =$ 37.78



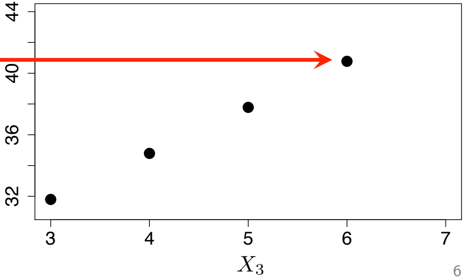
Constructing Partial Dependence Plots

Visually shows the effect of X_i on predictions *after accounting for other predictors*

	y	x1	x2	x3	yhat
1	35.70	2.17	1.77	6	34.84
2	52.28	2.42	5.63	6	51.84
3	38.18	0.78	2.74	6	41.46
4	35.99	0.09	3.04	6	43.91
5	21.19	2.21	0.50	6	29.07
6	54.38	-2.64	3.63	6	50.97
7	23.59	2.26	0.23	6	27.77
8	32.27	0.45	1.34	6	35.68
9	47.84	0.43	4.24	6	48.77
10	38.87	-0.84	2.60	6	43.42

$\hat{y}(X_3 = 6, X_{-3} = x_{-3,j})$

$\hat{y}(X_3 = 6) =$ 40.77



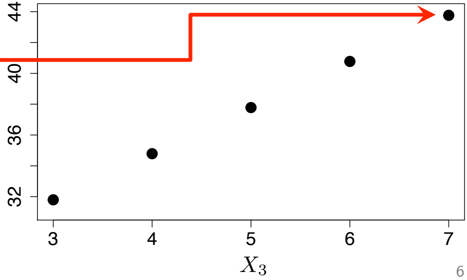
Constructing Partial Dependence Plots

Visually shows the effect of X_i on predictions *after accounting for other predictors*

	y	x1	x2	x3	yhat
1	35.70	2.17	1.77	7	37.84
2	52.28	2.42	5.63	7	54.83
3	38.18	0.78	2.74	7	44.45
4	35.99	0.09	3.04	7	46.90
5	21.19	2.21	0.50	7	32.07
6	54.38	-2.64	3.63	7	53.96
7	23.59	2.26	0.23	7	30.77
8	32.27	0.45	1.34	7	38.67
9	47.84	0.43	4.24	7	51.76
10	38.87	-0.84	2.60	7	46.41

$\hat{y}(X_3 = 7, X_{-3} = x_{-3,j})$

$\hat{y}(X_3 = 7) =$ 43.76



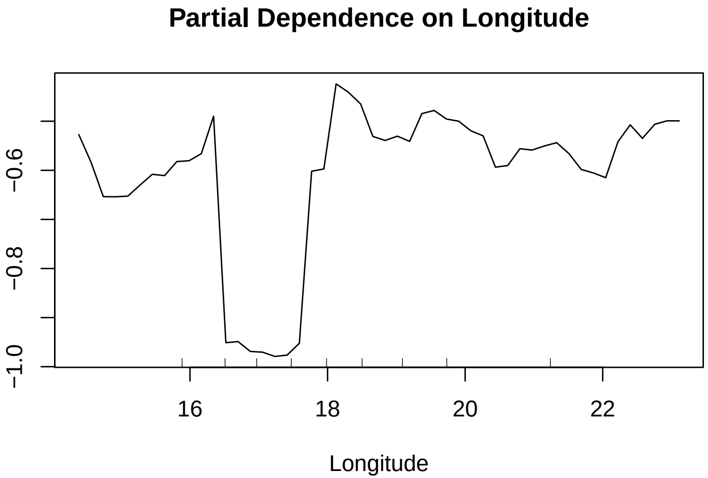
Partial Dependence Plots

Visually shows the effect of X_i on predictions *after accounting for other predictors*

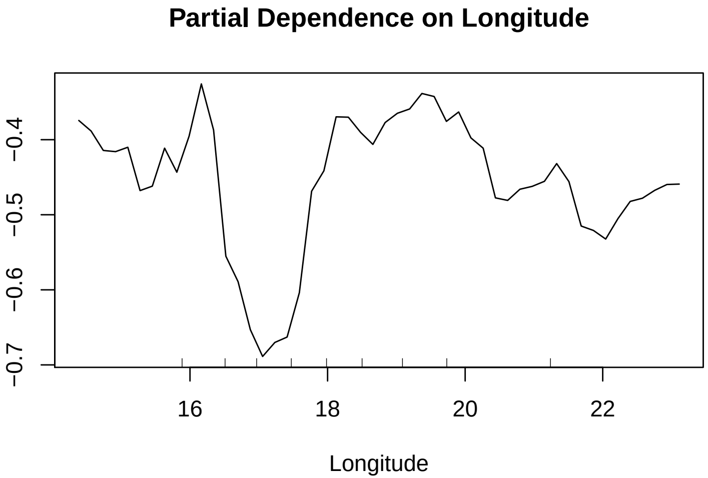
- Fix all sample data except for the data for X_i
- Replace all data for X_i with a small value, say x
- Get mean prediction $\hat{y}$
- Increase x by a small amount and repeat
- Plot all $(x, \hat{y})$ pairs

Note this is an *estimate* of the "true" partial dependency (since we use sample data)

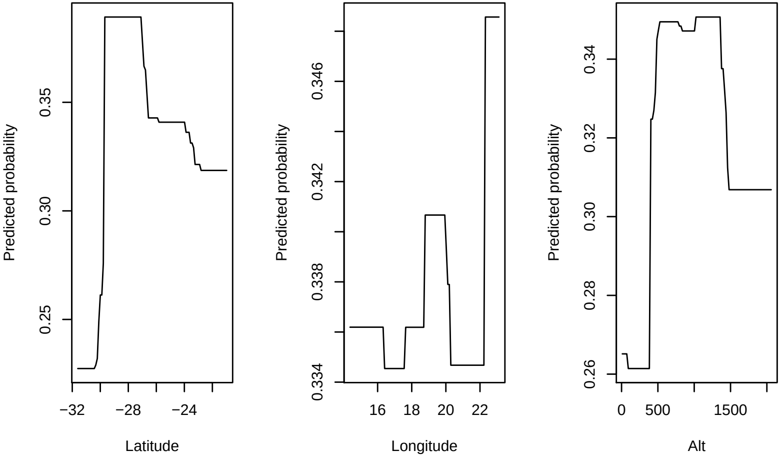
Partial Dependence Plots (bagging)



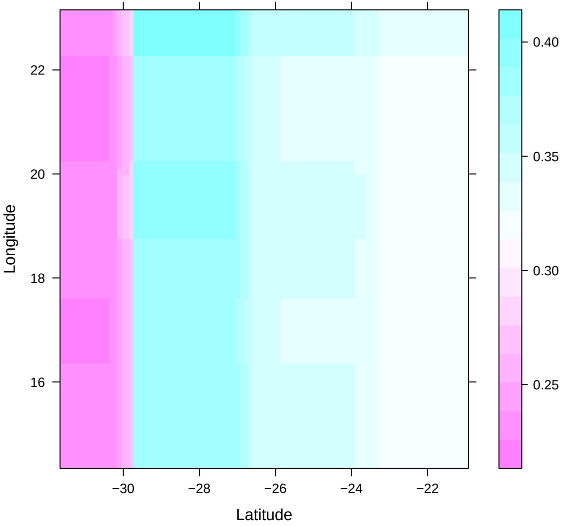
Partial Dependence Plots (random forest)



Partial Dependence Plots (boosting)



2-D Partial Dependence Plots (boosting)



Summary

- No traditional inference for ML methods
- Heuristic measures of variable importance
- Partial dependence plots show nature of relationships

- **Next:** Neural networks